AD_____

Award Number: W81XWH-10-1-1006

TITLE: Molecular Profiles for Lung Cancer Pathogenesis and Detection in U.S. Veterans

PRINCIPAL INVESTIGATOR: Steven M. Dubinett, M.D.

CONTRACTING ORGANIZATION: University of California, Los Angeles
Los Angeles, CA 90095

REPORT DATE: October 2012

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2012	Annual	20 September 2011 – 19 September 2012

4. TITLE AND SUBTITLE

Molecular Profiles for Lung Cancer Pathogensis and Detection in U.S. Veterans

5a. CONTRACT NUMBER

5b. GRANT NUMBER

W81XWH-10-1-1006

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Steven M. Dubinett, M.D. Pierre Massion, M.D.
Brigitte M. Gompers, M.D.
Ignacio Wistuba, M.D.
Avrum Spria, M.D.
E-Mail: sdubinett@mednet.ucla.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)
University of California, Los Angeles
Los Angeles, CA 90095-0001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
During our first year of research, we demonstrated a localized "field cancerization" phenomenon on gene expression in the airway of patients with lung cancer, and we identified several pathways preferentially activated in the airway adjacent to tumors. In addition, we have identified markers of stem cells in the airway that may represent tumor-initiating cells of the airway and are evaluating profiles of these cells. We have identified Snail as a novel marker of stem cells in the airway that promote epithelial-mesenchymal transition. We have made a major technical advance in developing the methods required to use low quality and quantity laser capture microdissected material to sequence the transcriptome. This allows us to examine the gene expression profiles in premalignant lesions and compare it to the histologically normal airway epithelium and tumor. We have validated this approach and the data will allow us to identify novel pathways for lung carcinogenesis. All of these studies are identifying biomarkers that could be used for early lung cancer detection and pathways that are involved in "field cancerization". Understanding this "field cancerization" and development of premalignant lesions is likely to shed light on novel pathways in lung carcinogenesis that could lead to diagnostic tests, therapies and cancer chemoprevention strategies for lung cancer.

15. SUBJECT TERMS
"field cancerization" ; lung carcinogenesis ; gene expression profiles ; lung cancer stem cells

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	70	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

INTRODUCTION

Lung cancer continues to be the leading cause of cancer-related death in both men and women in the United States [1]. The majority of lung cancers are non-small cell lung cancers (NSCLCs) that include squamous cell carcinomas (SCCs) and adenocarcinomas [2] Lung cancer mortality is high in part because most cancers are diagnosed after regional or distant spread of the disease had already occurred and due to the lack of reliable biomarkers for early detection and risk assessment [2]. The identification of new effective early biomarkers will improve clinical management of lung cancer and is linked to better understanding of the molecular events associated with the development and progression of the disease.

It has been suggested that histologically normal-appearing tissue adjacent to neoplastic lesions display molecular abnormalities some of which are in common with those in the tumors [3]. This phenomenon, termed field of cancerization, was later shown to be evident in various epithelial cell malignancies, including lung cancer [4,5]. Loss of heterozygosity (LOH) events are frequent in cells obtained from bronchial brushings of normal and abnormal lungs from patients undergoing diagnostic bronchoscopy and were detected in cells from the ipsilateral and contralateral lungs [6]. More recently, global mRNA expression profiles have been described in the normal-appearing bronchial epithelium of healthy smokers [7]. In addition, modulation of global gene expression in the normal epithelium in health smokers is similar in the large and small airways and the smoking-induced alterations are mirrored in the epithelia of the mainstem bronchus, buccal and nasal cavities [8].

In this program, in Specific Aim 1, high-throughput microarray mRNA expression analyses are being performed on cytological specimens (brushings) obtained at intraoperative bronchoscopy from the main carina and main ipsilateral bronchus, as well as on specimens obtained at lobectomy procedures from the main lobe bronchus (adjacent to SCCs), sub-segmental bronchus (adjacent to adenocarcinomas) and from the resected NSCLC tumors. Towards this aim, we are comparing and contrasting global gene expression patterns across all the specimens from the entire field and corresponding NSCLC tumors. We are deriving lung adenocarcinoma and SCC field cancerization signatures signifying the differential mRNA expression patterns between the carina and the subsegmental bronchus and main lobe bronchus, respectively. In addition, similar expression profiles between the carina and resected NSCLC tumors will be integrated with available gene expression data of bronchial brushings from the main carina isolated at various time points post-surgery from 40 NSCLC patients; Department of Defense (DoD) VITAL patients. Promising markers derived from this study are being validated at the mRNA and protein level in histological tissue specimens. Moreover, we are currently performing RNA-sequencing and microarray profiling of nasal epithelia, airway epithelial cells collected from both bronchoscopy and lobectomy specimens as well as of corresponding tumors (NSCLC patients) or benign lesions (cancer-free individuals).

In Specific Aim 2, we are using laser capture microdissection to obtain specific cell populations (basal cells or type II alveolar cells, depending on the NSCLC histology/location) as well as premalignant lesions and epithelial components of the tumors. These cell populations are being profiled with RNA-seq to determine their gene expression signatures to increase our understanding of premalignancy. We are analyzing the gene expression profiles that are associated with progression from a benign cell population to premalignancy and with progression from a benign cell population to true malignancy.

In future studies, in Specific Aim 3, we will use expression signatures and biomarkers derived from the results of aims 1 and 2 to develop and test airway-based biomarkers capable of diagnosing lung cancer in current or former smokers using minimally invasive sites.

This report details the progress made during the second year of research.

COMPREHENSIVE ANNUAL PROGRESS REPORT FROM LEAD PI (DR. STEVEN DUBINETT)

Molecular Profiles for Lung Cancer Pathogenesis and Detection in U.S. Veterans

Specific Aim 1: **To increase our understanding of the molecular basis of the pathogenesis of lung cancer in the "field cancerization" that develops in current and former smokers.**

Summary of Research Findings

A. Gene and protein expression analysis of bronchial epithelial samples obtained from bronchoscopy from NSCLC patients (Sub-specific Aim 1A, 1C and 1D): This analysis was performed on samples obtained from patients enrolled to the Vanguard clinical trial supported by the DoD VITAL grant (W81XWH-04-1-0412, PI Dr. W.K. Hong), and the gene profiling work and analysis, and the protein validation study, was partially supported by the Lung Cancer Consortium grant reported here (W81XWH-10-1-1007).

Gene Expression Analysis: Patients on the prospective Vanguard study had definitively treated ES (I/II) NSCLC and were current or former smokers. Patients had bronchoscopies with brushings obtained from the main carina (MC), airways adjacent (ADJ) to the previously resected tumor and from airways distant from the tumor in the ipsilateral (NON-ADJ) and contralateral (CONTRA) lung at baseline, 12, 24 and 36 months following resective surgery (**Figure 1**). Nineteen patients were selected for the study based on airway sampling of at least five different regions per time point and continuously up to 24 or 36 months (391 airway samples from nineteen patients). Data from the gene expression analysis have been described and detailed in the previous annual report (**Year 1).** Microarray analysis identified gene features (n=1165) that were non-uniform by site and differentially expressed between airways adjacent to tumors relative to more distant samples as well as those (n=1395) that were significantly altered with time up to three years. In addition, gene-interaction networks mediated by PI3K and ERK1/2 were modulated in adjacent compared to contralateral airways and the latter network with time. Key markers of the identified gene-interaction networks were validated at the protein level by immunohistochemical analysis. The study highlighted spatial and temporal cancer-associated expression alterations in the molecular field of injury of early stage NSCLC patients after definitive surgery that warrant further validation in independent studies. These findings were submitted and have been accepted for publication and are currently in press (see **Reportable Outcomes** and **Appendix**).

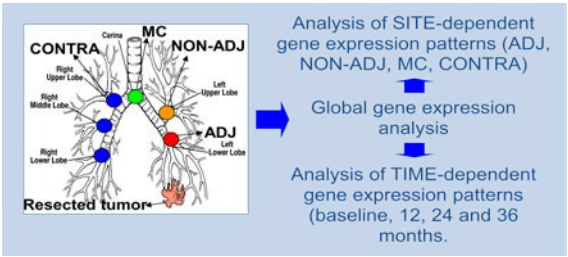


Figure 1. Schematic illustrating sampling and collection of bronchial brushes from different airway sites at different time points following resective surgery from early stage smoker NSCLC patients enrolled in the phase II cancer surveillance Vanguard study.

Protein Expression Analysis: Our findings on the modulation of PI3K- and ERK1/2-mediated networks by site and time after surgery prompted us to examine the immunohistochemical (IHC) expression of surrogate markers of both signaling cascades in corresponding formalin-fixed paraffin embedded (FFPE) airway biopsies. We sought to examine expression of AKT phosphorylated at Threonine308 since phosphorylation of this amino acid is well known to occur

through phosphoinositide-dependent kinase 1 (PDPK1) following PI3K activation. We assessed the immunohistochemical expression of phospho-AKT(Thr308) and phos pho-ERK1/2(Thr202/Tyr204) in available and ev aluable histologically normal bronchial epithelia biopsies (n=324) corresponding to the brushings analyzed by expression profiling. The intensity and extent of cytoplasmic and nuclear pAKT-Thr308 immunostaining was evaluated using a light microscope (magnification, x20). Cytoplasmic expression was quantified using a four-value intensity score (0, none; 1, weak; 2, moderate and 3, strong) and the percentage (0-100%) of the extent of reactivity). A final cytoplasmic expression score for pAKT-Thr308 was obtained by multiplying the intensity and reactivity extension values (range, 0-300). Nuclear expression was quantified using the percentage of extent of reactivity which gave rise to a nuclear expression score for pAKT-Thr308 (range, 0-100). Immunoreactivity of phospho-AKT (min 0, max 300) was variable as depicted in the representative photomicrographs in **Figure 2A** and was detected in both the cytoplasm and nucleus of normal bronchial epithelia (NBE) (**Figure 2A**). IHC analysis demonstrated that cytoplasmic ($p<0.0001$) and nuclear ($p=0.01$) phospho-AKT statistically significantly increased with time up to three years in all NBE (**Figure 2B**) with highest expression at the 36 month time point. Nuclear phosphorylated AKT was also statistically significantly increased in adjacent NBE compared to airways relatively more distant from tumors in the mixed-effects model ($p=0.04$) (**Figure 2B**). Immunoreactivity of phospho-ERK1/2 was also variable (min 0, max 209) and mainly cytoplasmic (**Figure 2C**). IHC analysis demonstrated that phosphorylated ERK1/2 was statistically significantly elevated in adjacent NBE ($p=0.03$) and significantly increased with time up to three years in all airways when averaged together ($p=0.02$) (**Figure 2D**) in the model. Notably, there was a significant interaction ($p=0.019$) between site of NBE and t ime of sampling as phospho-ERK1/2 expression was significantly increased in adjacent NBE but not in contralateral airways and main carinas in the model (**Figure 2D**) with highest expression in adjacent and non - adjacent (ipsilateral, green plot) airways. These data demonstrated that like differential gene expression profiles within the lung airway field of injury, canonical activated oncogenes are modulated by site from the resected tumor and time following definitive surgery in early stage NSCLC patients and have been also accepted for publication along with the gene expression data above (**Reportable Outcomes** and **Appendix**).

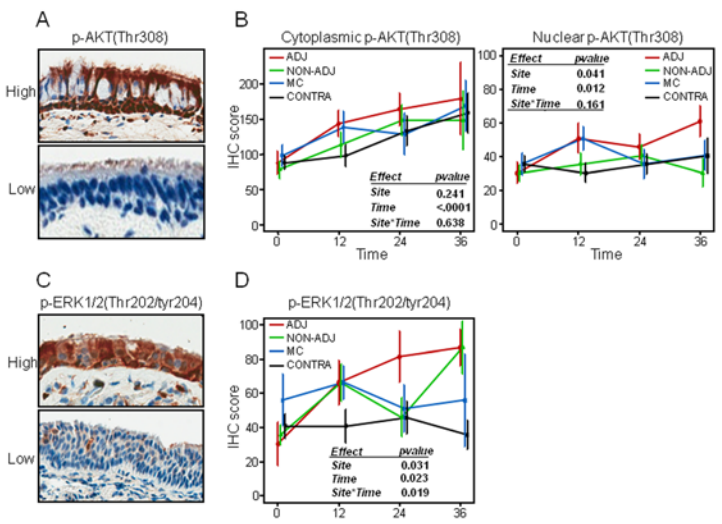


Figure 2. A. Representative photomicrographs (20x magnification) depicting strong (top) and weak (bottom) phospho-AKT(Thr308) immunostaining. **B.** Immunohistochemical scores of cytoplasmic (left) and nuc lear phospho-AKT (middle) were assessed for statistically significant differences by site and time in a m ixed-effect model and pl otted in main carinas (MC) and i n adjacent (ADJ), non-adjacent (ipsilateral, NON-ADJ) and contralateral (CONTRA) airways with time. **C.** Representative photomicrographs (20x magnification) depicting strong (top) and weak (bottom) phospho-ERK1/2(Thr202/Tyr204) immunostaining. **D.** Immunohistochemical scores of phosphorylated ERK1/2 levels were assessed for statistically significant differences by site and time in a mixed-effect model and repeated measures analysis, site*time, term for interaction between site and t ime factors. Error bars represent S.E. of the mean.

B. Gene expression analysis of bronchial epithelial samples obtained from lobectomy specimens from NSCLC patients (*Field Cancerization Study*) (Sub-specific Aim 1A and 1C):

Gene Expression Analysis: To increase our understanding of the molecular basis of lung cancer pathogenesis, we analyzed the transcriptome profiles of cytological specimens (brushings) obtained at lobectomy or pneumonectomy procedures from 2-5 bronchioles with differential proximity from resected tumors and from resected lung tumors and normal parenchyma (**Figure 3**). Samples were obtained from patients who did not receive neoadjuvant therapy undergoing lobectomy or pneumonectomy procedures (n=23) under an M D Anderson institutional review board (IRB)-approved protocol. A summary of the clinicopathological characteristics of the studied patient cases has been described in the previous annual report (**Year 1**). 194 samples were obtained from surgically resected specimens of early-stage (I-IIIA) NSCLC patients (n=20 cases). NSCLCs and paired normal lung were obtained frozen, preserved in RNA*later* or by surface-brushing. Normal bronchial epithelia (NBE) were collected by brushing bronchial structures using Cytosoft cytology brushes (Cardinal Health). Histological assessment of NBE by hematoxylin & eosin (H&E) and pan-cytokeratin IHC analysis confirmed the lack of preneoplastic and neoplastic cells and revealed greater than 90% epithelial content, respectively (**Figure 4**). Total RNA of all samples was isolated using the miRNeasy Mini kit following homogenization of tissues and brushing collections. RNA quality was assessed by the 28S/18S ribosomal RNA ratio and by RNA integrity number (RIN) using the Agilent Bioanalyzer. Processed RNA samples were hybridized to Affymetrix GeneChip® Human Gene 1.0 ST arrays and were then scanned using the GeneChip® Scanner 3000 from Affymetrix (Santa Clara, CA) to yield raw image files that were subsequently converted to probe set data. Raw data were normalized and processed using the robust multi-chip algorithm (RMA) algorithm. We then applied linear mixed-effects models and ANOVA-Tukey tests to identify differentially expressed genes among groups based on feature-by-featured t-test p-values and false discovery rates (FDRs).

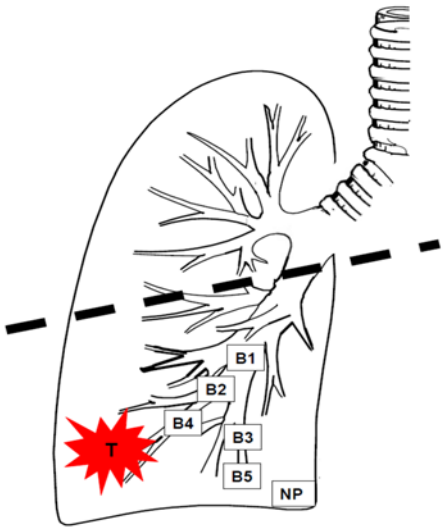


Figure 3. Schematic illustration of the lung airway for molecular mapping of field cancerization by high-throughput profiling. Tumor (T) and normal parenchyma (NP) samples were obtained by brushing, preservation in RNA*later* and snap-freezing in liquid nitrogen. Airway samples (B1 to B5) were obtained by brushing.

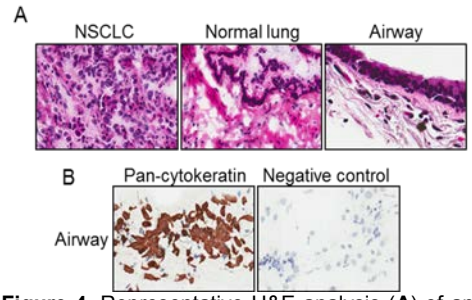


Figure 4. Representative H&E analysis (**A**) of an NSCLC (left), paired normal lung tissue (middle) and airway (right). **B**. Representative cytokeratin staining of an airway/bronchial structure showing positive immunoreactivity.

We then sought to derive gene expression signatures signifying genes significantly and concurrently differentially expressed between both tumors and airways compared to normal lung tissue (field cancerization signatures). We first analyzed tumors, airways and normal lung pertaining to NSCLC cases (n=20) excluding the three carcinoid cases. Following linear-mixed effects model coupled with ANOVA-Tukey analysis, we identified 457 and 1204 genes to be up-regulated and down-regulated, respectively, between both NSCLCs and airways compared to normal lung tissue based on a statistical threshold of a p-value<0.01 as well as a 1.5 fold-change in expression. Hierarchical clustering analysis demonstrated that these genes were able to separate NSCLC tumor samples and airway samples from normal lung tissues, the latter being segregated in a separate cluster (**Figure 5**). We then derived such field cancerization signatures when analyzing squamous cell carcinoma and lung adenocarcinoma cases (the two major subtypes of NSCLC) separately. Similar analysis was performed using the same selection criteria mentioned above. We derived field cancerization signatures denoting genes significantly and concordantly differentially expressed between both lung tumors and airways compared to paired uninvolved normal lung tissue for both lung SCC (n=5) and adenocarcinoma (n=14) cases that were comprised of 1,803 and 1,735 genes, respectively (**Figures 6A and 6B**). We then sought to functionally analyze the molecular field cancerization in NSCLC. We performed pathways and gene-interaction network analysis of the 1,661 differentially expressed genes described in **Figure 5** using Ingenuity Pathways Analysis (Redwood, CA). The top functional pathway and gene-interaction network are represented in **Figure 7A** and **Figure 7B**, respectively. Certain molecules within the top functional pathway (*MET*, *TGFBR2*) and gene-interaction network (*GPRC5A*, *VIPR1*) are highlighted in bold as they have been selected for further confirmation

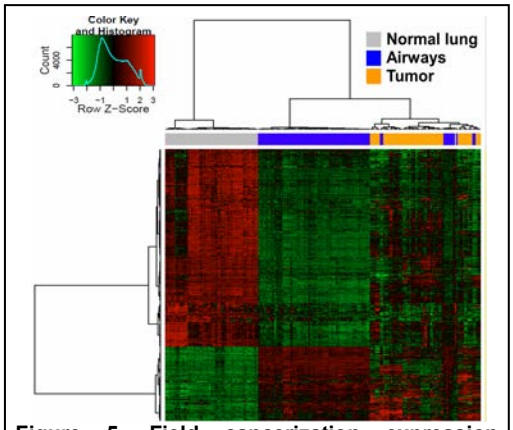


Figure 5. Field cancerization expression profiles in early-stage NSCLC. Hierarchical clustering of genes significantly differentially expressed between both tumors and airways compared to normal lung in NSCLC (n=1,661). Columns represent samples (n=194 samples from 20 cases) and rows represent gene features (red, up-regulated; green, down-regulated).

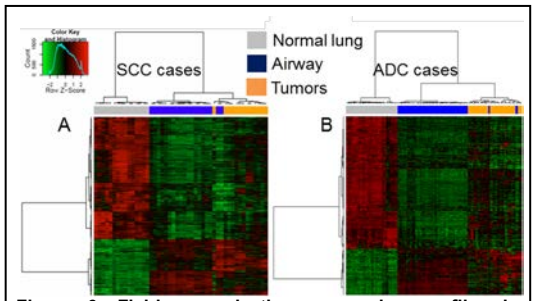


Figure 6. Field cancerization expression profiles in early-stage lung SCCs and adenocarcinomas. Hierarchical clustering of genes significantly differentially expressed between both tumors and airways compared to normal lung in SCC (**A**, n=1,803 genes) and adenocarcinoma (ADC) (**B**, n=1,735 genes) cases. Columns represent samples and rows represent gene features (red, up-regulated; green, down-regulated).

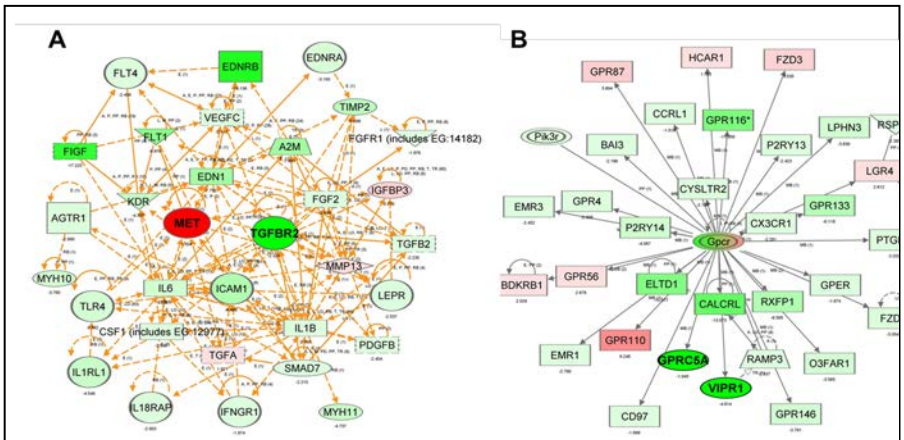


Figure 7. Functional pathways analysis of NSCLC molecular field cancerization. Functional pathway and gene-interaction network analysis of the NSCLC field cancerization signature derived in Figure 5. The top functional pathway and gene-interaction network are presented in **A** and **B**, respectively. Molecules selected for further expression validation are in bold. Red = upregulated genes, Green = downregulated genes.

and validation of expression in subsequent studies (see below). It is noteworthy that functional analysis revealed that the identified molecular field cancerization in NSCLC was comprised of typical cancer-associated pathways and networks such as embryonic stem cell, eicosanoid, retinoic acid receptor (RAR), NANOG stem cell, NF-kB and PTEN signaling pathways and gene networks. These findings highlight expression patterns and pathways that are typically deregulated in overt tumors but also prevalent in histologically normal airway epithelia, and therefore, may play important roles in the pathogenesis of lung adenocarcinomas and SCCs.

The above findings highlight expression patterns and pathways in the molecular field cancerization of NSCLC that may offer both detection markers as well as therapeutic targets to guide personalized chemoprevention in early stage NSCLC patients. To explore this concept, we assessed the relevance of the localized field cancerization expression profiles identified to those of large airways in lung cancer patients and high-risk smokers. We, therefore, analyzed the field cancerization profiles from Spira et al [9] comprised of 129 proximal airway samples. The gene expression raw data of airway epithelium of smokers with suspect lung cancer were re-normalized to produce gene-level expression values using RMA [10] in the affy package [11] of the Bioconductor software suite (version 2.4.1) [12] and Entrez gene-specific probe set mapping from BrainArray (version 11.0.1) [13]. A linear model was then used to adjust the expression estimates for the mean z-score quality metric previously described in the original report. We then performed Gene Set Enrichment Analysis (GSEA) [14] and a list of all Entrez gene identifiers within the field cancerization signature was ranked according to the Student *t* statistic computed between proximal airways from cancer and non-cancer patients from the original report by Spira et al. This list was then used to perform a pre-ranked GSEA analysis to identify which of the field cancerization gene sets were significantly enriched among genes that were up- or down-regulated in the normal bronchial epithelium of smokers with suspect lung cancer. A leading edge gene set was then derived that was comprised of 299 genes that were concordantly modulated between NSCLCs and airways compared to normal lung in the localized field cancerization as well as between proximal large airways of cancer patients compared to non-cancer patients (**Figure 8**). Side by side hierarchical cluster analysis using the leading edge genes of the localized field cancerization samples (left panel) and of the large airways of cancer and non-cancer patients (**right panel**) revealed concordance in differential expression among the sub groups (**Figure 8**). These data suggest that expression profiles from the localized field cancerization of resected NSCLC specimens may, in part, be relevant to differential field of injury molecular profiles between cancer and non-cancer patients and may comprise potential targets that can be used for both detection and chemoprevention strategies.

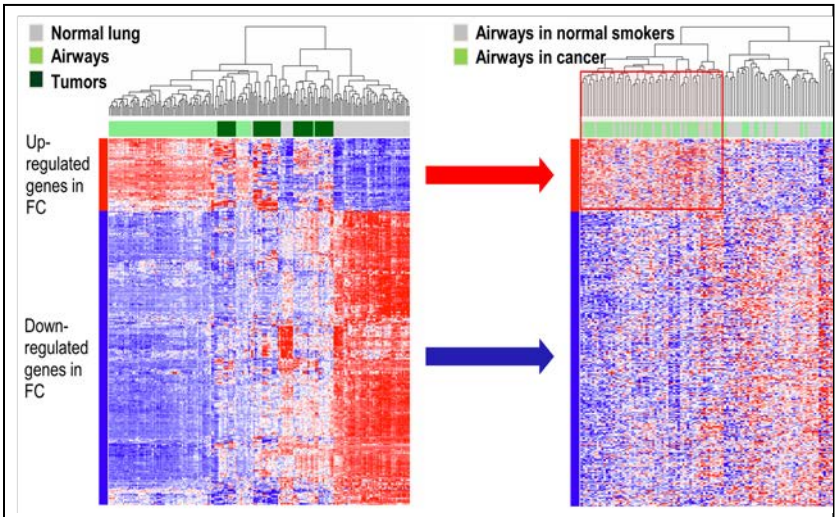


Figure 8. **Analysis of the localized field cancerization in uninvolved mainstem bronchi of patients with suspect lung cancer.** GSEA analysis was performed to identify genes (n=299) differentially expressed between tumors and airways compared to normal lung in the localized field cancerization (FC) of NSCLCs that are also significantly and concordantly differentially expressed between mainstem bronchi of cancer and non-cancer patients. Hierarchical cluster analysis using the 299 leading edge genes was performed side by side in the localized FC (left panel) and mainstem bronchi of cancer (green) and non-cancer (grey) patients (right panel).

We then sought to further study the transcriptomic architecture of the localized field cancerization to determine whether the molecular field cancerization exhibits site or distance from tumor-dependent expression patterns. Ordinal logistic analysis of airway samples alone identified genes significantly progressively and di fferentially expressed by distance from the tumors (n=422) (**Figure 9A**). Functional pathways analysis of the genes, using Ingenuity Pathways Analysis (Redwood, CA) demonstrated that canonical cancer-associated pathways, such as eukaryotic initiation factor, p70S6K kinase, polo-like kinase (PLK), gluconeogenesis and mammalian target of rapamycin (mTOR) signaling, were significantly and pr ogressively modulated in the airway epithelia by distance from the tumor (increased with shorter distance from tumors) (**Figure 9B**). We then derived a bioinformatic index score, signifying the extent of the site-dependent modulation of the genes in the airway cancerization, by computing the difference of expression values of the two clusters exhibited in the heatmap in **Figure 9A** across samples. The score/index quantitatively assessed the site-dependent effect and was increased with shorter distance of the airways from the tumors (**Figure 10A**). Notably, the extent of the site-dependent molecular field cancerization effect was more pronounced in lung SCC cases and i ncreased at a much larger rate with shorter distance from the tumors in this NSCLC subtype compared to adenocarcinomas (**Figures 10B and 10C**) which may be related in part to the proximal location of the sampled airways to SCCs relative to adenocarcinomas. However, it is noteworthy that we did not find significant differences in site-dependent expression of members of the surfactant family such as *SFTPB* and S*FTPC*.

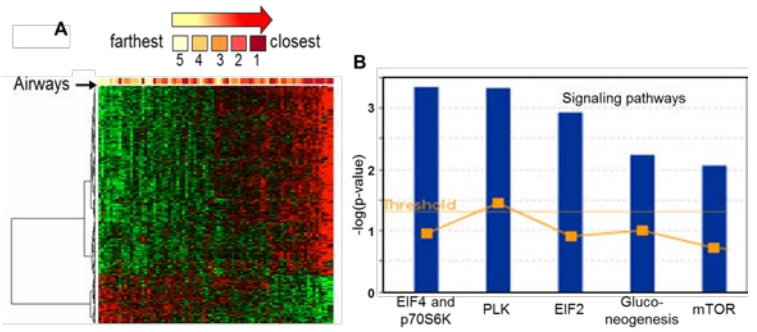


Figure 9. Site-dependent modulation of the molecular field cancerization in early-stage NSCLC. A. Ordinal logistic regression analysis of airways identifying 422 genes modulated in expression by distance from the tumor. **B**. Functional pathways analysis of the genes using Ingenuity Pathways Analysis software.

Since we performed the regression analysis on airways alone, we then deemed it worthy to determine whether the identified significant site-dependent gene expression modulation in the airways follow the same pattern between matched tumors and u ninvolved normal lung tissue. Importantly, the site-dependent genes signified by the cancerization index score were collectively significantly and concordantly modulated between tumors and pai red uninvolved normal lung tissues (**Figure 10D**). It is important to note that, although the site-dependent molecular field cancerization expression profiles were concordantly modulated between corresponding NSCLCs and matched normal lung tissue, they did not significantly correlate with expression profiles of the proximal large airways (mainstem bronchi) of patients with suspect lung cancer. GSEA analysis of the aforementioned 422 genes

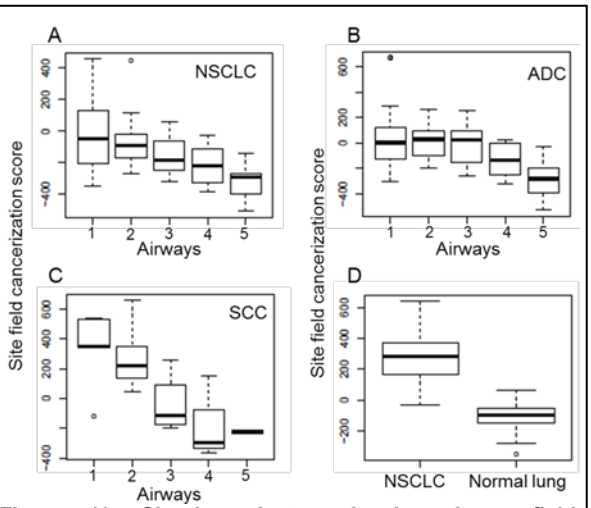


Figure 10. Site-dependent molecular airway field cancerization is more pronounced in lung SCCs relative to adenocarcinomas and informs of corresponding tumor profiles. Boxplots depicting site-dependent molecular airway field cancerization modulation in early-stage NSCLCs (**A**), adenocarcinomas (**B**), lung SCCs (**C**) and bet ween paired NSCLCs and normal lung tissue (**D**). Airway distance from tumor is numerically indicated with a range of 1 (closest) to 5 (farthest).

(clustered in **Figure 9**) did not reveal significant enrichment of this expression signature in the published dataset by Spira et al [9]. These findings suggest that unlike global localized field cancerization signatures we described above, site (from tumor)-dependent expression differences in the localized molecular field cancerization may not be relevant to the extended airway field of injury. Nonetheless, it is plausible to assume that the site-dependent localized molecular field cancerization profiles comprise expression profiles with typical cancer-associated signaling pathways and inform of tumor expression patterns and, thus, of potential early mechanisms of lung cancer pathogenesis commencing in normal airway epithelia.

Expression validation studies: We then sought to validate, in part, expression of field cancerization markers. We selected field cancerization markers that displayed site-dependent or –independent differential expression following microarray profiling analysis of the localized field cancerization based on both statistical criteria and functional pathways analysis. Out of the NSCLC cases studies, 18 comprised samples with enough RNA left after array profiling for performing quantitative real-time polymerase chain reaction (QRTPCR) analysis. Microarray profiling identified lysosome associated protein transmembrane 4 beta (*LAPTM4B*) and neuropilin (NRP) and tolloid (TLL)-like 2 (*NETO2*) to exhibit site-dependent differential expression (increasing in airways closer to corresponding tumors) in the localized field cancerization. Both genes were not identified in the global field cancerization signatures derived (**Figures 5 and 6**) when all samples were analyzed independent of site. Trans*forming growth factor receptor beta II (TGFBR2)* was found to be decreased in the molecular field cancerization in a site-independent manner and *vasoactive*

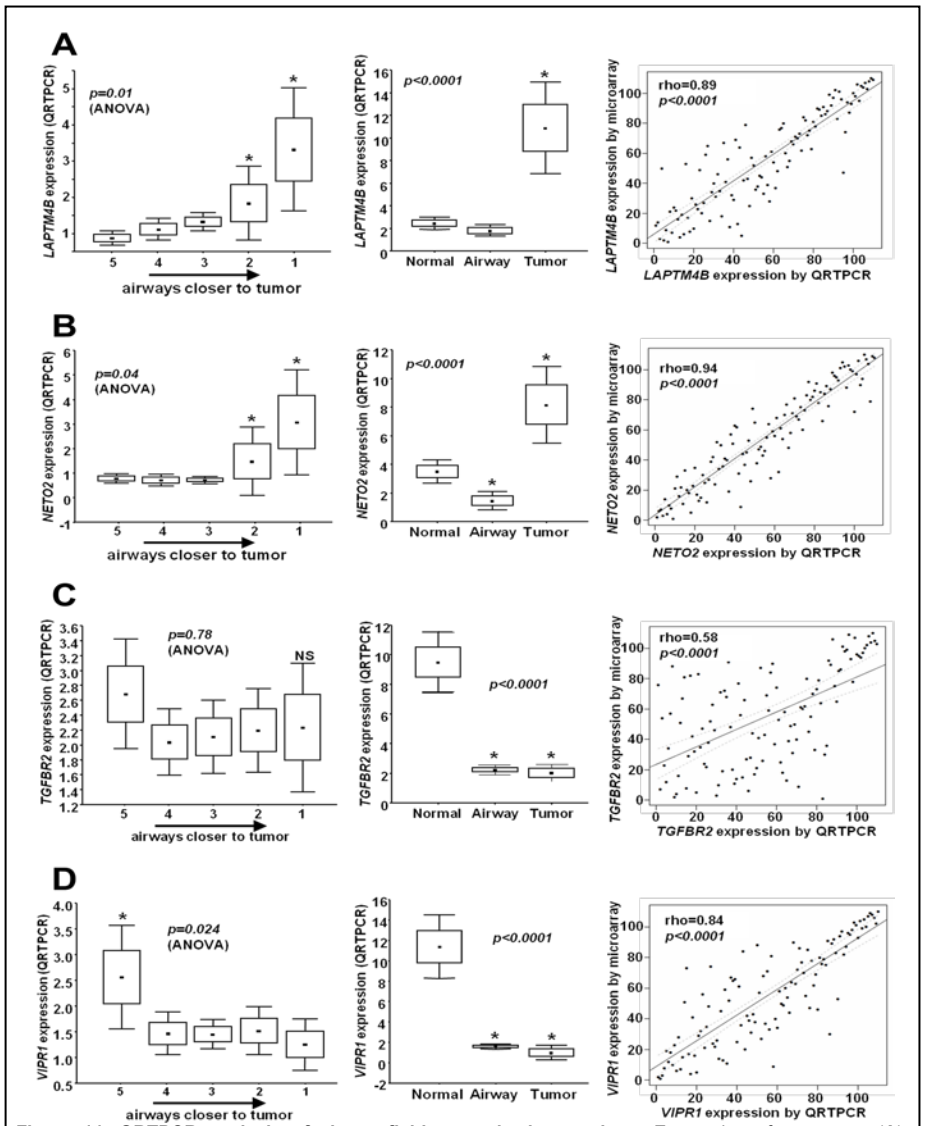


Figure 11. QRTPCR analysis of airway field cancerization markers. Expression of *LAPTM4B* (**A**), *NETO2* (**B**), *TGFBR2* (**C**) and *VIPR1* (**D**) was analyzed by QRTPCR in 18 out of 20 cases with sufficient RNA from samples. QRTPCR analysis was performed in a site-dependent manner (left panels) and in a site-independent manner comparing tumors, airways and normal lung tissues (middle panels). Correlation between expression quantified by microarray and QRTPCR analyses was statistically assessed by Spearman rank (right panels).

intestinal peptide receptor 1 (VIPR1) exhibited both site-dependent and independent decreased expression in the molecular field cancerization. We performed QRTPCR analysis of *LAPTM4B*, *NETO2*, *TGFBR2* and *VIPR1* on all evaluable cases (**Figures 11 A-D**). QRTPCR analysis confirmed the microarray findings of the differential expression pattern of all four genes in the molecular field cancerization (**Figures 11A-D**). *LAPTM4B* and *NETO2* were significantly increased in airways closer to tumors compared to more distant normal epithelia ($p<0.05$) (**Figures 11A-B, left panels**). Similar to the array findings, *LAPTM4B* and *NETO2*, although increased in tumors compared to normal lung tissue, exhibited similar levels of expression among airways, when averaged all together, and normal lung (**Figures 11A-B, middle panels**). Moreover and closely similar to the microarray data, *TGFBR2* and *VIPR1* mRNA expression were decreased in tumors and airways compared to normal lung tissue (**Figures 11C-D, middle panels**). Whereas *TGFBR2* did not exhibit site-dependent differential expression patterns (**Figure 11C, left panel),** *VIPR1* was progressively decreased in airways with closer distance from corresponding tumors (**Figure 11D, left panel**). Importantly, we found significant and high correlation between expression of the four genes analyzed by microarray and expression quantified by QRTPCR (all $p<0.0001$) (**Figures 11A-D, right panels**). We next performed western bot analysis of the 20 NSCLC field cancerization case tumors and corresponding uninvolved normal lung tissues for assessing expression of the protein products coding for *NETO2* and *TGFBR2*. Western blot analysis confirmed the tumor versus normal differential expression of the field cancerization markers at the protein level (**Figure 12**). It is noteworthy that we were not successful to assess the protein expression of LAPTM4B as the limited antibodies available for this recently identified gene did not work well in our hands.We also confirmed expression of additional field cancerization markers. Microarray analysis revealed that the *MET* oncogene was increased in the molecular field cancerization, and did not exhibit site-dependent differential expression features. We analyzed *MET* mRNA and protein expression by QRTPCR and immunohistochemical analysis, respectively, in an independent set of cases and samples we had not profiled (n=6 cases). QRTPCR validated the microarray findings in the independent set and demonstrated that *MET* mRNA expression was significantly higher in both lung tumors and airways compared to normal lung tissue (*p<0.0001*) (**Figures 13A-B**). Immunohistochemical analysis confirmed the array

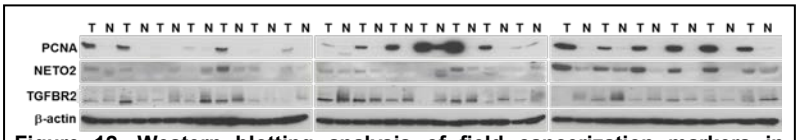


Figure 12. Western blotting analysis of field cancerization markers in NSCLCs and corresponding normal lung tissues. NSCLC and matched normal lung tissues (20 field cancerization case pairs) were lysed and homogenized after which 20 μg of the total protein homogenates were subjected to SDS-PAGE. Membranes were incubated with primary antibodies raised against NETO2 and TGFBR2. Membranes were also incubated with antibody against proliferating cell nuclear antigen (PCNA) to confirm tumor and normal samples and with β-actin to ensure equal loading.

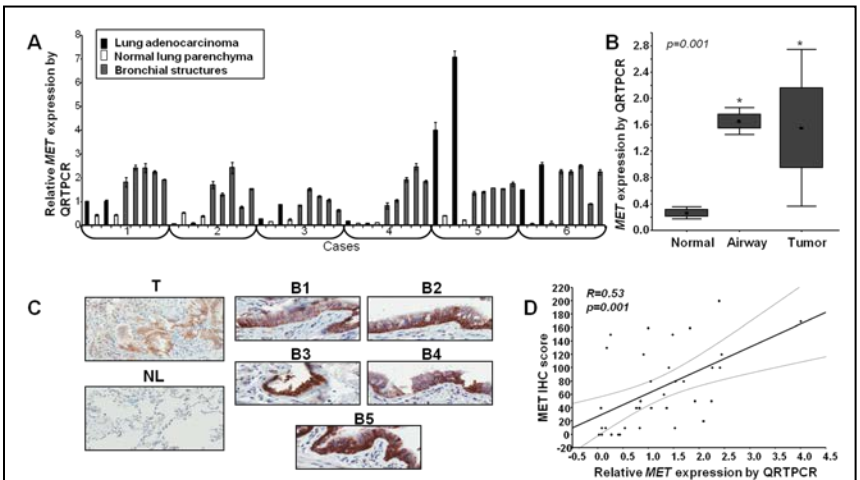


Figure 13. Validation of increased expression of the *MET* oncogene in the field cancerization. Quantitative real-time PCR analysis of *MET* mRNA expression in adenocarcinomas and matched airways and normal lung (**A and B**). **C**. Representative MET immunohistochemical expression in a case comprised of a tumor, uninvolved normal lung and five bronchial structures. **D**. Correlation plot of *MET* mRNA and protein expression in the same cases by Pearson correlation.

and QRTPCR findings at the protein level and demonstrated increased total MET immunohistochemical expression in tumor and normal bronchial epithelial histological tissue sections compared to those of normal lung tissue (**Figure 13C**). In addition and importantly, we found statistically significant correlation ($p<0.001$) between *MET* mRNA assessed by QRTPCR and protein expression analyzed by IHC in the same samples studied (**Figure 13D**).

Our microarray analysis also pinpointed to a decreased expression of the lung-specific tumor suppressor gene G-protein coupled receptor family C, group 5, member A (*GPRC5A*) in tumors and airways compared to uninvolved normal lung tissue. We performed QRTPCR analysis of *GPRC5A* mRNA expression in NSCLCs, normal lung tissue and bronchial epithelia from six independent cases we had not profiled. QRTPCR analysis confirmed the microarray findings and demonstrated that *GPRC5A* expression was significantly decreased in the molecular localized field cancerization (**Figures 14A-B**). *GPRC5A* was identified as a lung-specific tumor suppressor gene evidenced by spontaneous adenocarcinoma formation in mice with knockout of both of the gene's alleles [15]. More recently, *GPRC5A* was shown to exert its tumor suppressive function, in part, by inhibition of nuclear factor-kappa B (NFκB) and downstream inflammation [16]. Therefore, we sought to analyze *GPRC5A* expression in the molecular field cancerization associated with chronic obstructive pulmonary disease (COPD), a risk factor for lung cancer that is typically associated with inflammation [17]. GPRC5A expression was assessed in histological tissue specimens of NSCLC and normal bronchial epithelia including those from lung cancer-free COPD and from NSCLC patients. Representative photomicrographs of GPRC5A immunohistochemical staining in the different normal bronchial epithelia is depicted in **Figure 15A**. GPRC5A immunoreactivity appeared to be highest in normal bronchial epithelia from lung disease-free never- and ever-smokers and successively significantly decreased in NBE of COPD patients who are cancer-free and lowest in NBE of patients with both COPD and adenocarcinoma or SCC (**Figure 15B**). A general linear model

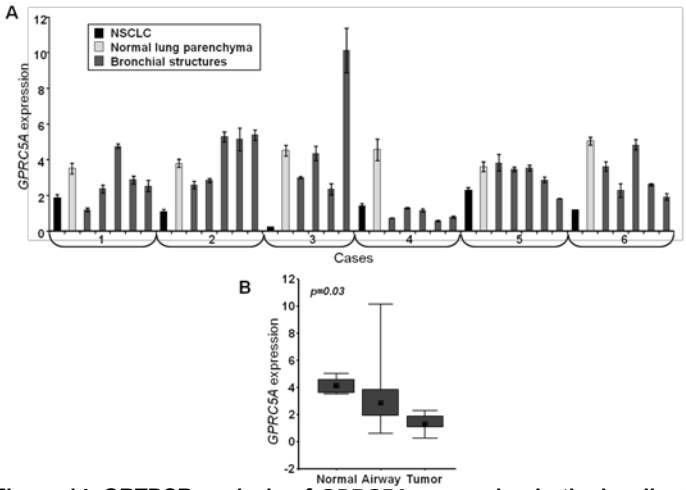


Figure 14. QRTPCR analysis of *GPRC5A* expression in the localized field cancerization. Total RNA isolated from tumors, normal lung tissue and normal bronchial epithelia from six NSCLC cases (A) was analyzed by QRTPCR for GPRC5A expression. **B**. Box plot showing average GPRC5a expression in NSCLCs, bronchial epithelia and normal lung tissue statistically analyzed by the Kruskal-Wallis test.

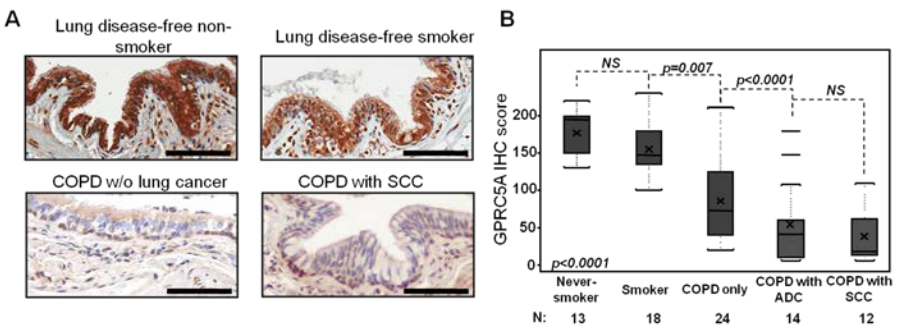


Figure 15. GPRC5A expression in COPD-associated field cancerization. **A**. Representative photomicrographs of GPRC5A expression in NBE from disease-free never- and ever-smokers (upper left and right, respectively) and in NBE from cancer-free COPD patients (bottom left) and NSCLC patients with COPD (bottom right). **B**. Box-plots depicting quantification of GPRC5A immunohistochemical expression represented in (**A**). Main P-value signifies statistical significance of reduced GPRC5A expression in the sequential pathogenesis of COPD-associated NSCLC and among the five groups by the general linear model and Kruskal-Wallis test.

demonstrated a significant gradual decrease of GPRC5A expression from NBE of disease-free never-smokers to NBE from patients with COPD and adenocarcinoma or SCC ($p<0.0001$) (**Figure 15B**). The mean (177.31 ± 30.66) and median (195; min, 130; max, 220) of GPRC5A expression in NBE of lung disease-free never-smokers were higher than that of smokers (mean, 155.28 ± 34.15, median, 147.5; min, 100; max, 230) although the difference was not statistically significant (**Figure 15B**). Moreover, GPRC5A expression was significantly higher in NBE of lung disease-free smokers relative to NBE in cancer-free COPD patients ($p=0.007$), which in turn was significantly higher compared to the expression in NBE from COPD patients with adenocarcinoma ($p<0.0001$) or SCC ($p<0.0001$) (**Figure 15B**). Furthermore, although GPRC5A expression was lower in NBE of COPD patients with SCC compared to those with COPD and adenocarcinoma, the difference was not statistically significant (**Figure 15B**). These findings suggest that reduced expression of the GPRC5A tumor suppressor may be implicated in the pathogenesis of NSCLC associated with inflammatory COPD. Our findings on the differential expression of GPRC5A in the molecular field cancerization including that associated with COPD have been accepted for publication and are in press (see **Reportable Outcomes** and **Appendix**).

Functional studies: As mentioned above, we confirmed and validated the differential expression of selected field cancerization markers in matched NSCLCs, bronchial epithelia and uninvolved normal lung tissue. Some of the field cancerization markers exhibit well-reported roles in lung cancer pathogenesis (e.g., *MET*, *GPRC5A*) whereas other validated molecules are either recently shown to function in lung carcinogenesis [*TGFBR2* [18,19]] or have not been reported to exhibit aberrant function in the pathogenesis of lung malignancies (e.g., *LAPTM4B* and *NETO2*). We were prompted to assess the expression and potential functional relevance of the latter group of under studied genes in an *in vitro* model of lung carcinogenesis comprised of normal (NHBE), immortalized (BEAS-2B and 1799) and transformed (1198) and tumorigenic (1170) bronchial and lung epithelial cells [19].

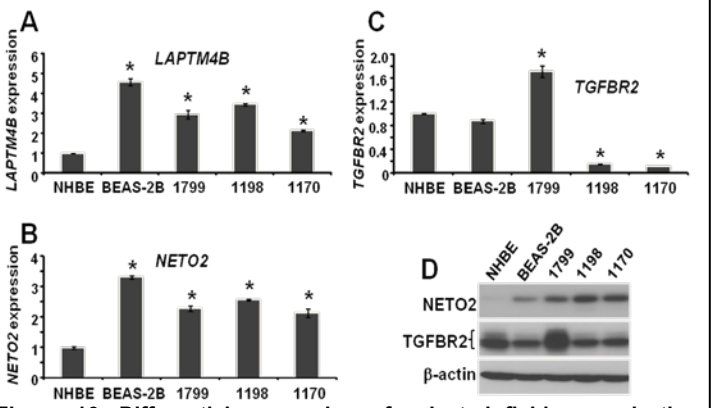


Figure 16. Differential expression of selected field cancerization markers in an *in vitro* model of lung carcinogenesis. QRTPCR analysis of *LAPTM4B* (**A**), *NETO2* (**B**) and *TGFBR2* (**C**) expression in normal (NHBE), immortalized (BEAS-2B and 1799) and transformed and tumorigenic (1198 and 1170) lung epithelial cells. D. Western blot analysis of NETO2 and TGFBR2 protein expression in the in vitro model. Membranes were stained with antibody against actin to ensure equal loading.

QRTPCR analysis demonstrated that, in accordance with our findings on differential expression patterns in the molecular field cancerization, *LAPTM4B* and *NETO2* were both up-regulated in tumorigenic lung epithelial cells compared to their normal counterparts whereas *TGFBR2* was down-regulated in the tumorigenic cells (**Figures 16A-C**). Moreover, the differential expression of *NETO2* and *TGFBR2* in the utilized *in vitro* model was confirmed at the protein level by western blot analysis (**Figure 16D**). These findings demonstrate that novel field cancerization markers may have important relevance to lung cancer pathogenesis.

Microarray analysis of the field cancerization in early-stage NSCLC revealed that *LAPTM4B*, a lysosome associate transmembrane putative oncogene with no known role in lung carcinogenesis [20], was among the top 5 site-dependent field cancerization markers.

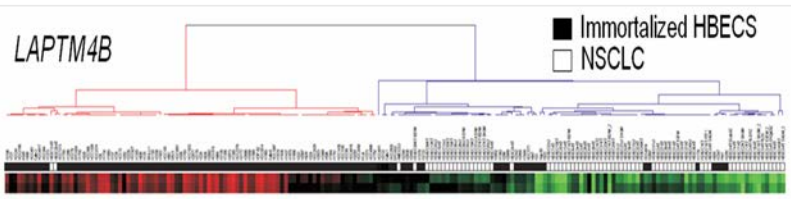


Figure 17. *LAPTM4B* expression in NSCLC and immortalized bronchial epithelial cell lines. Increased expression of *LAPTM4B* in NSCLC cells compared to HBECS by microarray analysis.

Moreover and in our hands, *LAPTM4B* was significantly elevated in NSCLC and immortalized bronchial epithelial cell lines compared to normal cells. We then explored *LAPTM4B* expression in a large available array dataset of NSCLC cell lines from Dr. John D. Minna (University of Texas Southwestern, Dallas, Texas). Microarray analysis of the cell lines demonstrated that *LAPTM4B* expression was significantly higher (left cluster) in a fraction of NSCLC cell lines compared to bronchial cells (right dendogram cluster) (**Figure 17**). These findings utilizing different cell culture systems suggest that field cancerization markers, exemplified by *LAPTM4B*, may be modulated in expression in the pathogenesis sequence of lung cancer.

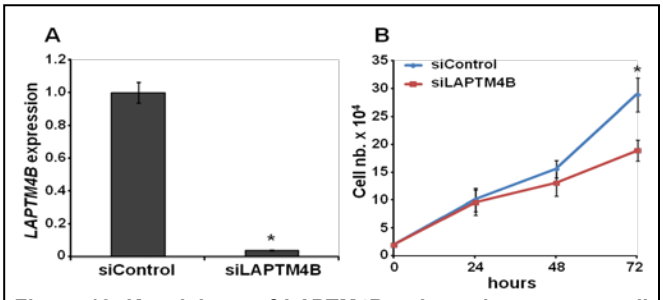


Figure 18. Knockdown of LAPTM4B reduces lung cancer cell growth. A. QRTPCR analysis of LAPTM4B expression in Calu6 cells transfected with control siRNA or *LAPTM4B*-specific siRNA. **B**. Trypan blue exclusion count of cells at 24, 48 and 72 h following transfection of control or *LAPTM4B*-targeting siRNA.

We then attempted to preliminarily assess the functional relevance of the field cancerization marker *LAPTM4B* to NSCLC malignant phenotype. Microarray analysis of the large array of NSCLC cell lines (**Figure 17**) aided us to select cell lines with relatively high expression of the gene. We employed SMARTpool small interfering RNA (Dharmacon, Inc., Lafayette, CO) that consist of pooled siRNAs targeting different regions of the *LAPTM4B* transcript. Transient transfection of Calu-6 NSCLC cells with *LAPTM4B*-targeting siRNA effectively reduced *LAPTM4B* expression (**Figure 18A**) and significantly decreased cell growth evidenced by trypan blue exclusion cell count (**Figure 18B**). The effect of RNA interference-mediated knockdown of LAPTM4B *expression on cell growth was* confirmed in another NSCLC cell line (H1650, **Figures 19A, C and E**) and in the immortalized bronchial epithelial cell line BEAS-2B (**Figures 19 B, D and F**) further demonstrating that LAPTM4B is a positive mediator of cell growth in NSCLC cell lines. We sought to further confirm the role of *LAPTM4B* in lung cancer cell growth by

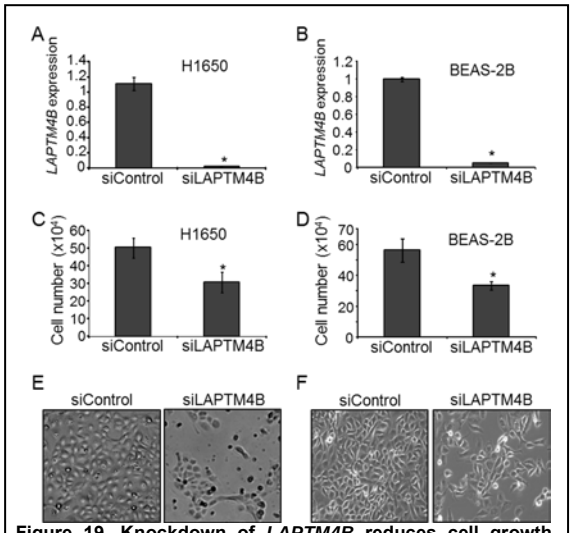


Figure 19. Knockdown of *LAPTM4B* reduces cell growth. QRTPCR (**A-B**) and cell growth analysis by trypan blue exclusion (**C-D**) of NSCLC and immortalized bronchial epithelial cells transfected with control or *LAPTM4B* specific siRNA. Phase-contrast micrographs (x20) of control and *LAPTM4B*-siRNA transfected H1650 (**E**) and BEAS-2B (**F**) cells.

generating sub-lines with stable knockdown of the gene in the Calu-6 cell line with high basal level of *LAPTM4B*. E mpty vectors and vectors containing scrambled small hairpin RNA (shRNA) as well as shRNA specific to *LAPTM4B* (Origene, Rockville, MD) were used for generation of the stable sub-lines. Stable selection was performed by incubating transfected cells for two weeks in cell culture medium with 10 μg puromycin. QRTPCR demonstrated significantly reduced *LAPTM4B* expression in Calu-6 cells with stable knockdown of *LAPTM4B* compared to cells stably transfected with scrambled shRNA or empty vectors (**Figure 20A**). Moreover, Calu-6 cells with stable knockdown of *LAPTM4B* displayed significantly reduced cell growth and anchorage-dependent colony formation (**Figure 20B** and **Figures 20C-D**, respectively). These data demonstrate that *LAPTM4B* field cancerization marker is a positive mediator of immortalized and malignant lung epithelial cell growth. Moreover, these findings are proof of principle and demonstrate that functional characterization of field cancerization markers in the lung airway can lead to the identification of otherwise unknown functional targets and mediators of lung cancer molecular pathogenesis. Our findings on the aforementioned microarray analysis of the field cancerization as well as on the expression validation and functional studies of field cancerization markers have been presented as an abstract and are currently being prepared as a manuscript for publication which we anticipate submitting for peer-review by the end of the year (see **Reportable Outcomes** and **Appendix**).

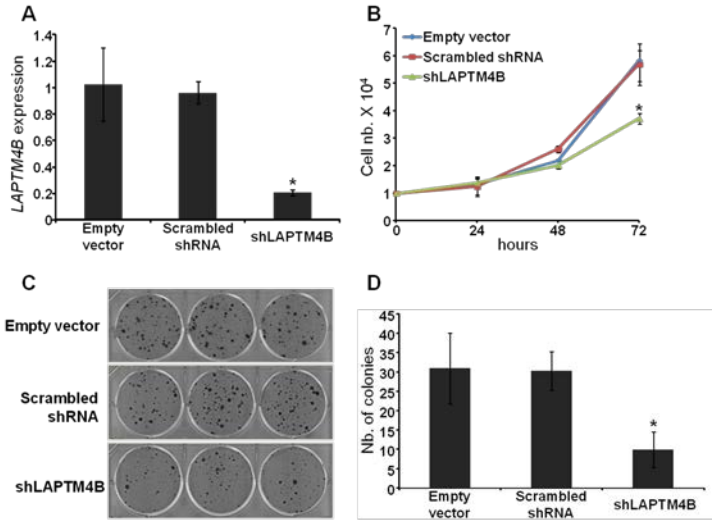


Figure 20. Stable knockdown of LAPTM4B reduces lung cancer cell growth and colony formation. A. QRTPCR analysis of Calu-6 cells stably transfected with empty vectors or vectors containing scrambled shRNA or *LAPTM4B*-specific shRNA. **B**. Trypan blue exclusion method count of the stably transfected cells. **C-D**. Anchorage-dependent colony formation assay of the stably transfected Calu-6 cells was performed by seeding cells at a seeding density of 500 cells per well in six-well plates. Representative triplicate images were obtained with phase-light microscope (**C**) and quantified to assess number of colonies (**D**).

C. Collection of epithelial samples from both bronchoscopy and lobectomy specimens from patients with lung cancer (Sub-specific Aims 1A and 1C):

We obtained consent from 35 study participants to collect tissue samples. From these subjects we have collected sets of bronchoscopy (nasal and bronchial brushes) and lobectomy (tumor, normal parenchyma, and bronchial brushes) epithelial samples (**Table 1**).

Table 1. Specimen types collected

Specimen type	Number of specimens
Blood	26
Bronchial epithelial brushings	145
Distal airway brushings	3
Nasal epithelial brushings	33
Tumor tissue	8

The samples are currently being analyzed by both next generation RNA-sequencing (RNA-Seq) using the Illumina HiSeq 2000 platform and microarray profiling using the Human Gene 2.0 ST platform from Affymetrix. RNA-Seq and microarray analysis are being performed at BU and MD Anderson Cancer Center, respectively. The study was uniquely designed by collecting tumor, normal lung tissue and airway samples from cases with NSCLC obtained from all four participating institutions (**Table 2**). Samples from cases without lung cancer were collected and processed from BU, UCLA and Vanderbilt (**Table 3**). The demographics of the patients are shown in **Table 4**.

Total RNA from all samples in the different institutions were isolated similarly using the miRNeasy kit from Qiagen according to the manufacturer's instructions. Hiseq RNA sequencing will facilitate the discovery of novel transcripts in the molecular field of injury as well quantifying expression of those that could not be characterized by microarray technology. In addition and in contrast to the Human Gene 1.0 ST microarray platform used for the field cancerization profiling described above, the newer 2.0 ST platform to be utilized will enable the interrogation of the complete set of known functionally important long intergenic non-coding RNAs (lincs) in the molecular field of injury [21]. To date this is the first study that is 1) performing next generation sequencing in addition to microarray profiling analysis of the molecular filed of injury; 2) applying multiple high throughput technologies on field of injury samples obtained from four different institutions in the nation and 3) cross-comparing the complete topological map of the extended and localized molecular field of injury/cancerization between both NSCLC patients and cancer-free individuals. We anticipate that RNA-Seq and microarray profiling will be completed by the end of the year with subsequent bioinformatic and functional analysis along with validation of expression studies completed by Spring 2013. Due to the use of both next generation RNA sequencing and comprehensive microarray profiling and due to this ongoing study's unique design we anticipate that expression profiles in the NSCLC molecular field of injury will harbor molecules, both novel and established, that may exhibit potential for use as airway biomarkers that can be developed and tested for lung cancer detection using minimally invasive sites in Specific Aim 3 of this award.

Table 2. Molecular mapping of the field of injury in NSCLC and cancer-free patients

	RNA-Seq (cases)			Microarray (cases)		
Institution	**ADC**	**SCC**	**No Cancer**	**ADC**	**SCC**	**No Cancer**
MD Anderson	4	2	0	4	3	0
BU	2	1	3	0	2	4
UCLA	2	2	1	3	2	2
Vanderbilt	1	1	1	4	3	1
Number	9	6	5	11	10	7
Total Nb. of cases analyzed	**20**			**28**		
Total Nb. of samples analyzed	**156**			**183**		

RNA-Seq, RNA sequencing; ADC, adenocarcinoma; SCC, squamous cell carcinoma; BU, Boston University; UCLA, University of California Los Angeles.

Table 3. Diagnoses of patients participating in our study

Number of Subjects	Diagnosis
14	Lung cancer
5	Benign
1	Other cancer
15	Pending

Table 4. Demographics of study participants

Ethnicity	Male	Female
White	14	5
Black	7	3
Hispanic	1	0
Asian	2	1
American Indian	0	0
Other	1	0
Unknown	0	0

Specific Aim 2: To increase our understanding of the role of tumor-initiating stem/progenitor cells in the pathogenesis of lung cancer in the "field cancerization" that develops in current and former smokers.

Summary of Research Findings:

A. Feasibility of sequencing small amounts of RNA from laser captured samples that reflect different pathologic stages of lung carcinogenesis (Sub-specific Aim 2B):

Specific regions of normal basal cells, premalignant metaplastic/dysplastic cells, and squamous carcinoma cells were successfully selected by laser microdissection. This was described and detailed in the previous annual report (Year 1). Adequate amounts of RNA were isolated from these cells for library preparation and high throughput sequencing (RNA-seq), and dec ent quantities of libraries with appropriate size ranges were generated. This was described and detailed in the previous annual report (Year 1). Samples were sequenced on Illumina Genome Analyzer IIx or HiSeq 2000 instruments, producing single-end reads with quality control Phred scores above 30.

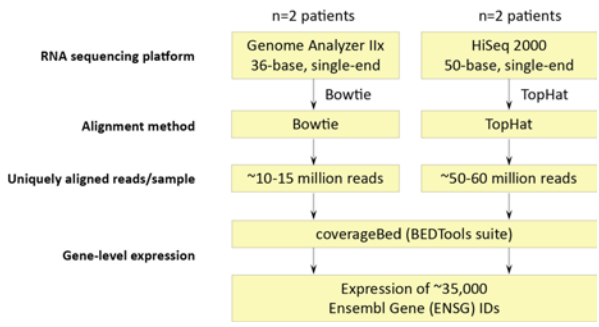


Figure 21. R NA sequencing workflow. Single-end reads were aligned to the human genome and gene ex pression was quantified by measuring the overlap with Ensembl Gene IDs.

Sequence alignment and quantification of gene expression

The reads produced by sequencing each RNA sample were aligned to the human genome (build hg19). Uniquely aligned reads were then used to compute gene expression estimates by measuring the coverage of each of ~35,000 Ensembl Gene loci using the coverageBed utility in the BEDTools suite. This workflow is illustrated in **Figure 21**.

Identification of genes with progression-associated expression patterns

To identify genes whose expression was associated with progression from normal to intermediate (metaplastic or dysplastic) to tumor cells, a three-step procedure was used as outlined in **Figure 22**. First, genes with low expression (median expression was below 100 reads), which are more likely to be false positives, were removed from analysis. Next, to simplify the analysis, a concordance filter was then applied in order to consider only those genes whose expression changed in the same direction in both intermediate and tumor cells relative to normal cells.

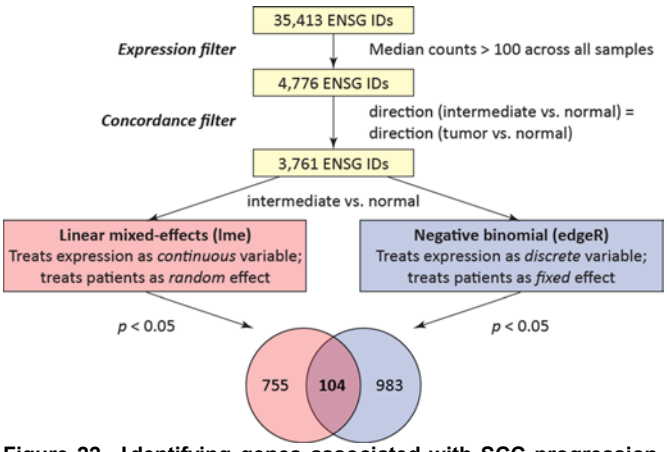


Figure 22. Identifying genes associated with SCC progression. Genes were filtered by median expression and by direction of change relative to normal controls before assigning significance using two

Finally, two different statistical models were used to identify genes whose expression was significantly altered between intermediate and normal cells. In one case, a linear mixed-effects approach was used, which assumes that expression is a continuous measure, and in which sample group and patient were modeled as fixed and random effects, respectively. Separately, a negative binomial model was used, which assumes that expression is a discrete variable, and in which both patient and sample group were modeled as fixed effects. The gene lists obtained with a nominal *p* value cutoff of 0.05 by either approach were intersected to obtain a list of 104 genes. The expression of these genes in all samples from all four patients is illustrated in **Figure 23**.

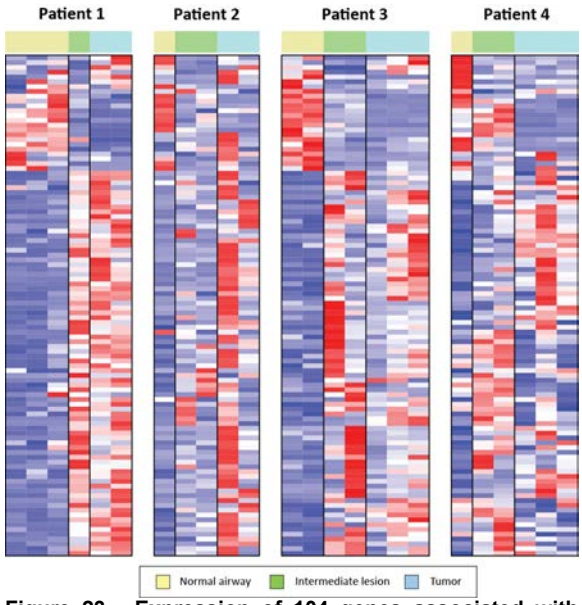


Figure 23. Expression of 104 genes associated with SCC progression. The raw counts have been scaled to a mean of zero and unit variance within each patient. Blue and red indicate expression that is two or more standard deviations below or above the mean expression in each row, respectively.

B. Data Validation with Quantitative Real-Time PCR and Immunofluorescence Staining:

To validate the data from gene expression analysis of the first four patients, we performed quantitative real-time PCR (qPCR) on selected

candidate genes. From the RNA-seq data, we calculated the coefficient of variance of the reads per kilobase per million reads (RPKM) across all samples for 23 genes that were commonly used as endogenous control genes in TaqMan-based qPCR assays. From this analysis, we chose Beta-2-microglobulin (B2M) as the endogenous control, as it has the lowest coefficient of variance as well as the highest median read counts. We used two of the sequenced patients, Patients 3 and 4, as well as four additional independent cases, Patients 5–8, for the qPCR validation study. We selected three candidate genes for validation, CEACAM5, SLC2A1, and PTBP3, as all of them showed increased expression in the premalignant lesions and tumors compared to normal basal cells. For each of these genes, we examined the changes of mRNA levels between intermediate lesions and the normal basal cells across all six patients (**Figure 24A**). The mRNA levels of CEACAM5 and SLC2A1 were validated to be higher in the intermediate lesions compared to normal basal cells across all patients. For PTBP3, two of the six patients examined showed lower mRNA levels in the intermediate lesions than normal basal cells.

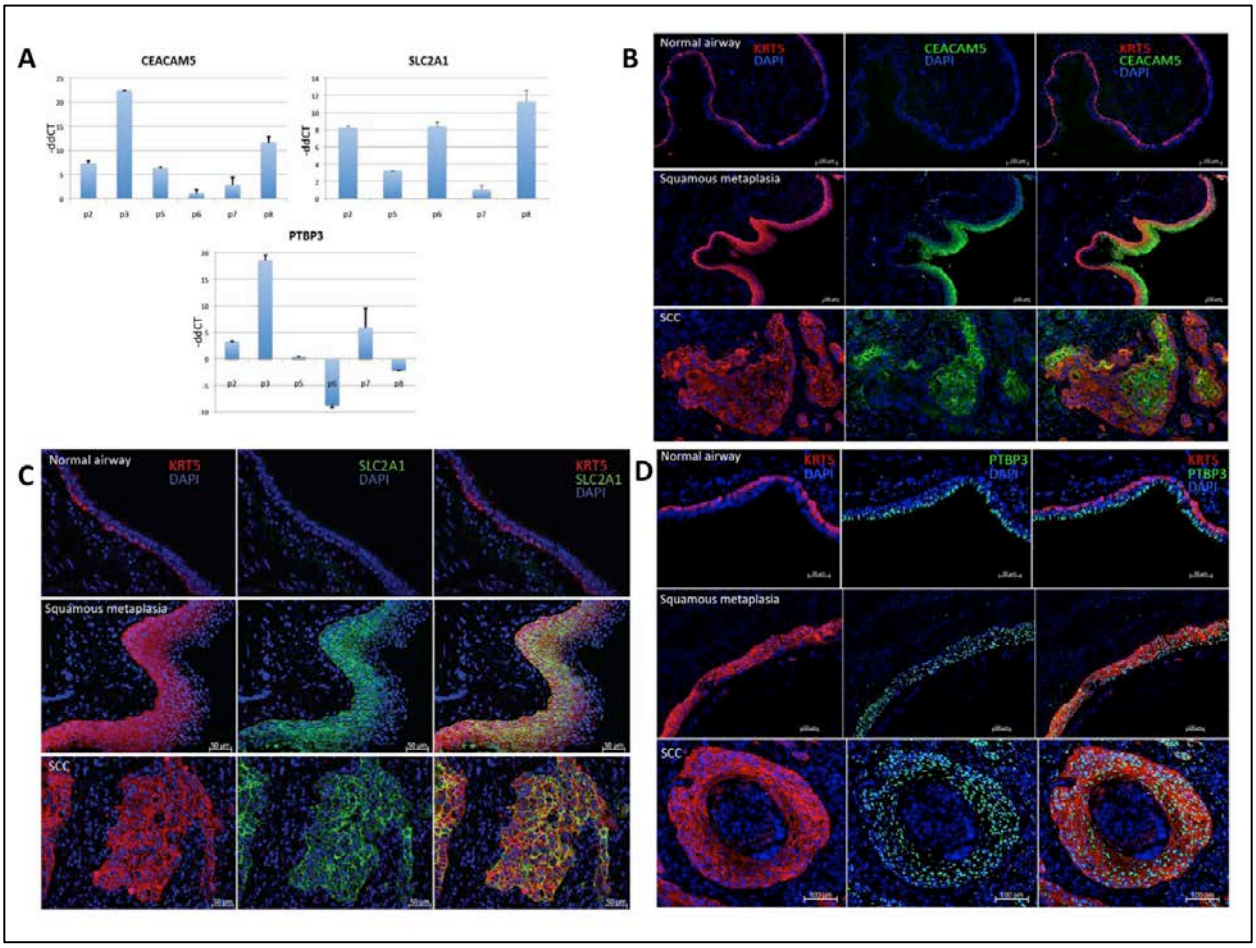


Figure 24. Validation of RNA-seq data at the mRNA and protein level.
A. Q-PCR with Taqman primer probe sets showed increased gene expression of SLC2A1 and CEACAM5 in premalignant lesions and tumors as compared to normal basal cells. PTBP3 expression was similarly increased, except in 2 patients. **B-D**. Immunofluorescent staining for CEACAM5, SLC2A1 and PTBP3, respectively. All 3 proteins were expressed at higher levels in premalignant lesions and NSCLCs as compared to normal basal cells. Scale bars = 100 microns. Blue = DAPI staining. KRT5 = keratin5 , a basal cell marker.

We picked these 3 genes for validation because they all could play critical roles in the biology of premalignant lesions and lung carcinogenesis. SLC2A1 is also known as glucose transporter 1 (GLUT-1) and is a facilitative glucose transporter. This isoform may be responsible for constitutive or basal glucose uptake and has broad substrate specificity. It is critical in glucose metabolism, and driving glycolysis in particular. CEACAM5, a cell surface glycoprotein that plays a role in cell adhesion and in intracellular signaling, is part of the CEACAM family that is important in other epithelial cell cancers, such as colon cancer. PTBP3 functions as an RNA-binding protein that mediates pre-mRNA alternative splicing regulation. It plays a role in the regulation of cell proliferation, differentiation and migration and is a positive regulator of erythropoietin-dependent erythropoiesis.

It has been reported that the correlation between mRNA levels and protein levels is not always straightforward.[22-24] To further investigate if the changes in transcript levels in the intermediate lesions for the selected genes do carry on to the protein levels, we performed tissue immunofluorescence staining of SLC2A1, CEACAM5, and PTBP3 on tissues with normal epithelium, premalignant lesions, and carcinomas from two independent cases (Patients 9 & 10) (**Figure 24B–D**). Since the PTBP3 data were not validated in Patient 6 and Patient 8 with qPCR, additional immunofluorescence staining of PTBP3 was performed on these two patients to examine if the data could be validated at the protein level. We found that CEACAM5 and SLC2A1 were not expressed in the normal epithelium, but were highly expressed in premalignant lesions. SLC2A1 was also highly expressed in the tumor, erefore validating the RNA-seq data at the protein level. Similarly, we found that CEACAM5 was expressed at a low level in the normal airway basal cells, had increased protein expression in the premalignant lesions and higher expression in tumors. Unlike SLC2A1, CEACAM5 was expressed in some but not all tumor cells. PTBP3 was expressed in the columnar cells of normal epithelium but showed low to no expression in the basal cells. PTBP3 expression was also detected in premalignant lesions and in tumor cells and some mesenchymal cells surrounding the tumor.

We also examined the relationship between smoking history and the expression of SLC2A1 and CEACAM5 in the airway epithelium in a previous gene expression profiling study of bronchoscopic brushings (**Figure 25**) [Spira et al PNAS 2004]. The expression of both genes was strongly increased in brushings from current smokers compared with those from never smokers (p=8.5x10^{-8} and 2.2x10^{-14}, respectively). However, the two genes differed in their reversibility after smoking cessation. SLC2A1 expression was not significantly different between former and never smokers (p=0.20), but was significantly different between current and former smokers (p=0.012). Conversely, CEACAM5 expression was not significantly different between current and former smokers (p=0.061), but was still significantly induced in former smokers relative to never smokers (p=9.9x10^{-5}). These data indicate that SLC2A1 and CEACAM5 are reversibly and irreversibly induced by smoking, respectively.

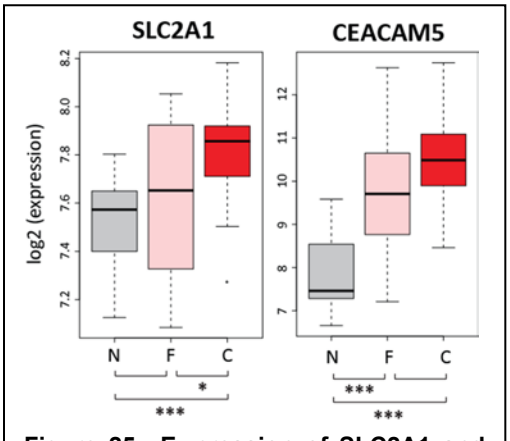


Figure 25. Expression of SLC2A1 and CEACAM5 expression in the bronchial epithelium is associated with smoking status. Gene expression was profiled by microarray in bronchial brushings from never (N), former (F) and current (C) smokers. [Spira et al PNAS 2004] * p < 0.05 by Student's t test; *** p < 0.001.

In this study, we have used laser microdissection to obtain materials from the tumors, premalignant lesions, and normal cells of the same patient, which offers a more accurate depiction of differential gene expression than in previously published lung cancer studies of

undissected lung tissue sections.[25-27] These data are currently the subject of a manuscript in preparation.

C. Identification of additional archived clinical specimens for laser microdissection of tumor-initiating stem/progenitor cells:

Additional archived clinical specimens from which we previously extracted DNA and RNA from rare target cells using laser microdissection were identified. Within the same individual, the following regions/cells were present: normal epithelium, normal basal cells, dysplastic cells in preinvasive tissue, as well as their respective basal cells, and invasive tumor cells. The cases identified included biopsy specimens and resected tumors, and each was confirmed to be amenable for microdissection after review of freshly cut H&E stained slides by our collaborating pathologists. Using archived clinical specimens available in biorepositories at each site, we have sufficient material to complete the studies described in sub-specific Aim 2B.

D. Feasibility of performing shotgun proteomics on rare target cells and arriving at a robust protein signature (Sub-specific Aim 2B):

As described in the previous annual report (Year 1), we have identified a molecular profile dominated by the Snail transcription factor that appears to drive epithelial mesenchymal transition (EMT) and tumor-initiating characteristics in the airway epithelium, as modeled *in vitro* and *in vivo*. Snail is also over-expressed in human bronchial epithelial cells in premalignant lesions *in situ* concomitant with markers of EMT and stemness. To validate the technology for analyzing tumor-initiating stem cells from *in situ* specimens, we performed preliminary Shotgun Proteomic Analysis comparing *in vitro* human bronchial epithelial cells and the same cells ectopically over-expressing Snail. Cell pellets were collected, prepared for LC-MS/MS shotgun proteomics, and analyzed, as described in detail in our year 1 progress report. Briefly, each 0.2 mg protein aliquot was digested and resolved by isoelectric focusing into 15 fractions that were subsequently analyzed by LC-MS/MS. Thus, there were 6 measurements (2 technical replicates for 3 samples) for the control group and 6 for the Snail+ group. Raw MS/MS data were evaluated using MyriMatch and IDPicker software. Differentially expressed proteins were then identified using Quasi-Tel pair wise comparison. The initial dataset was robust, with 2809 protein groups identified overall; a protein group usually represents a single protein, but it is occasionally a small group of indistinguishable proteins with identical peptides. The overall numbers of protein groups in the control and Snail+ bronchial epithelial cells were similar (2229 and 2738, respectively). The following general observations were made: (1) Known markers of EMT were over-expressed in the Snail+ cells. (2) Other structural/motility proteins consistent with an EMT phenotype were also over-expressed in the Snail+ cells.

To augment our ability to identify proteins relevant to the molecular pathogenesis of lung cancer across the broadest possible patient population, we will perform shotgun proteomics on additional samples. A panel of bronchial epithelial cells isolated from patients and engineered to over-express Snail have been plated in western blot and anchorage independent growth (AIG) assays. Via these assays, we will re-confirm their Snail expression and Snail-driven malignant conversion prior to their proteomic evaluation. Each of the cell types in this panel has previously demonstrated numerous Snail-driven cancer-associated phenotypes, including EMT, stemness, AIG, and/or tumor growth and metastatic behavior in mice. At the conclusion of these assays, cells maintained in culture in parallel will be collected, prepared for shotgun proteomics, and analyzed, as previously described. Additionally, tumor-initiating Snail+ALDH+CD44+CD24- bronchial epithelial cells will be subjected to shotgun proteomics in the same manner. This relatively rare cell type will be isolated by fluorescence activated cell sorting, and the resulting

cell pellets will be frozen. Multiple pellets will be pooled to generate material sufficient for evaluation; this again models laser microdissection of rare stem/progenitor target cells from archived clinical specimens.

By evaluating additional samples, including rare tumor-initiating stem cells, we anticipate arriving at a more robust protein signature relevant to lung carcinogenesis. Models and software developed in the Jim Ayers Institute at Vanderbilt, in collaboration with Dr. Dan Liebler, are more appropriately applied to studies with these multiple inputs. The new protein signature that emerges will be further strengthened via Multiple Reaction Monitoring (MRM) performed on the remaining samples. MRM using mass spectrometry is a highly sensitive and selective method for the targeted quantitation of protein or peptide abundances in complex samples. While shotgun proteomics detects all protein changes in the sample in an unfocused fashion, MRM is targeted and highly selective, allowing us to specifically look for proteins of interest. Dr. Liebler has developed and refined a series of MRM assays that can be applied to the airways of individuals for lung cancer. The relevant method developments were recently published in MCP [28]. To this end, we have generated a list of candidate proteins for MRM utilizing shotgun proteomic, mRNA array, and miRNA array datasets generated from the same Snail+ cells. Candidates with the greatest fold change and level of significance were included. Candidates at the intersection of each of these lists, as evaluated by Ingenuity Pathway Analysis (IPA), were also included. Finally, additional candidates of interest were included based on our hypothesis-driven studies of lung carcinogenesis, including mediators of inflammation, EMT, stemness, metabolism, apoptosis-resistance, as evaluated in the PIs' lab-based studies over the past several years. While we have already generated this candidate list for MRM, the list will be further refined as we expand our shotgun proteomic analysis to include additional samples.

Finally, during the preceding funding period, we developed a Microsoft Access database with the intent of including an additional parameter, "druggability", in the selection of top candidates for further validation and detailed functional studies. The first iteration of the database was created by integrating the protein, mRNA, and miRNA datasets previously described with a lung cancer-specific terms list. These were then linked to information regarding proteins/genes/miRNAs for which agents are in use or in the pipeline along with additional clinical utility parameters, such as how successful the agent is and its range of use. This database will be refined as we expand our shotgun proteomic analysis to additional samples and as we receive inputs from the sequencing and array studies in the other aims. This database will serve as an important new tool for selecting the best protein candidates to include in our upcoming MRM studies.

Specific Aim 3: Test airway-based mRNA and microRNA biomarkers of diagnosing lung cancer in current and former smokers at high risk for lung cancer in minimally invasive sites.

Summary of Research Findings:
The studies on this Aim will be carried out in Years 3 and 4 of the grant.

KEY RESEARCH ACCOMPLISHMENTS

- Identified aberrant activation of canonical oncogenes in the molecular field of injury of early stage NSCLC patients including phosphorylated AKT and ERK kinases.
- Completed the analysis and characterization of the molecular localized field cancerization using 194 samples comprised of tumor, normal lung and airway samples from 20 NSCLC cases.

- Derived field cancerization expression signatures, comprised of genes concordantly and significantly differentially expressed between tumors and ai rways compared to matched normal lung tissue, and pertinent to NSCLC, lung adenocarcinomas and SCCs.
- In collaboration with BU (Partnering institution, demonstrated that molecular profiles in the localized field cancerization are, at least in part, relevant to the molecular field of injury and contain markers significantly and concordantly different between airways of patients with and without lung cancer
- Identified profiles that were significantly progressively and di fferentially expressed by distance from corresponding lung tumors and concordantly modulated between tumors and paired uninvolved normal lung tissues, pinpointing to their probable roles in pathogenesis.
- Confirmed the differential expression of several site-dependent and –independent markers identified by microarray analysis of the field cancerization at the RNA and protein level by QRTPCR and immunohistochemical analysis, respectively.
- Identified previously unknown promalignant roles for *LAPTM4B*, one of the top 5 site-dependent field cancerization markers we had found, in NSCLC evidenced by reduced clonogenic growth of cells with knockdown of the gene.
- Began to analyze by RNA-Seq (20 cases, 156 samples) and comprehensive microarray profiling (28 cases, 183 samples) the molecular field of injury in NSCLC patients and cancer-free individuals.
- Performed RNA sequencing on c ell populations in matched sets of histologically normal airway, premalignant lesions and tumors from the same individuals, and identified candidate genes that increase in expression in premalignancy and in tumors.
- Identified and validated candidate genes and their proteins that are likely to play a role in the biology of premalignancy and lung carcinogenesis including SLC2A1 (glucose transporter), CEACAM5 (cell adhesion molecule) and PTBP3 (regulator of cell proliferation, differentiation and migration).

REPORTABLE OUTCOMES

Abstracts:

- Kadara H, Fujimoto J, Yoo SY, Garcia M, Kabbout M, Basey A, Wang J, Coombes KR, Kim ES, Hong WK, Kalhor N, Moran C, Wistuba II. Gene expression profiling of lung tumors and matched normal airways reveals common and disparate aberrant pathways in squamous cell carcinoma and adenocarcinoma development and potential targets for chemoprevention in early stage lung cancer patients. Proceedings of the 103rd Annual Meeting of the American Association for Cancer Research; 2012 Mar 31- Apr 4; Chicago, Illinois. Philadelphia (PA): AACR; 2012. Abstract #1721.
- Ooi AT, Gower AC, Zhang KX, Vick J, Caballero N, Massion PP, Wistuba II, Walser TC, Dubinett SM, Pellegrini M, Lenburg ME, Spira A and Gomperts BN. Molecular Profiles to Improve our Understanding of Lung Cancer Pathogenesis in U.S. Veterans. NIH Lung Cancer SPORE Meeting. Pittsburgh. July 2012.

Manuscripts:

- Kadara H, Wistuba II. Field cancerization in NSCLC: implications in disease pathogenesis. *Proceedings of the American Thoracic Society*, 9(2):38-42 2012.
- Fujimoto J*, Kadara H* (* equal contributing first authors), Garcia MM, Kabbout M, Behrens C, Liu DD, Lee JJ, Solis LM, Kim ES, Kalhor N, Moran C, Shalafkhaneh A, Lotan R, Wistuba II. G-protein coupled receptor family C group 5 member A (GPRC5A) expression is decreased in the adjacent field and normal bronchial epithelia of patients with chronic

obstructive pulmonary disease and non -small cell lung cancer. *Journal of Thoracic Oncology*. In Press.

- Kadara H, Shen L, Fujimoto J, Saintigny P, Chow CW, Lang W, Chu Z, Garcia M, Kabbout M, Fan YH, Behrens C, Liu D, Mao L, Lee JJ, Gold KA, Wang J, Coombes K, Kim ES, Hong WK, Wistuba II. Characterizing the molecular spatial and temporal field of injury in early stage smoker non-small cell lung cancer patients after definitive surgery by expression profiling. *Cancer Prevention Research*. In Press.

CONCLUSIONS

During our second year of research, we derived expression profiles signifying the molecular field cancerization in early-stage NSCLC. We also showed that the field cancerization profiles, identified in distal airways of resected specimens, are, at least in part, relevant to the extended field of injury and differentially expressed between large airways of patients with and without lung cancer. In addition, our profiling studies revealed expression patterns in bronchial epithelia that are significantly progressively modulated with distance from corresponding tumors and, importantly, concordantly modulated between tumors and paired uninvolved normal lung tissues, pinpointing their probable roles in pathogenesis. In addition, based on functional pathways analysis and statistical threshold criteria, we validated the differentially expression of several key site-dependent and-independent field cancerization markers. Furthermore, we found that, *LAPTM4B*, a putative oncogene with no known role in lung carcinogenesis and one of top 5 site-dependent field cancerization markers that increased with shorter distance from tumors in our profiling studies, was a positive mediator of the malignant phenotype. We also used a unique approach to profile cell populations from the normal airway, premalignant lesions and tumors and were able to validate these genes. We identified SLC2A1, CEACAM5 and PTBP3 as genes that are upregulated in premalignancy and squamous lung cancer. In particular, the glucose transporter SLC2A1 represents a novel potential target for preventing or treating premalignant lesions and, potentially, lung cancer. Current efforts are underway to profile and perform RNA-Seq on the complete topological map of the field of injury in both NSCLC patients and cancer-free individuals, which is expected to yield airway biomarkers for lung cancer to be tested in future aims of the study.

REFERENCES

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69-90.
2. Herbst RS, Heymach JV, Lippman SM. Lung cancer. N Engl J Med 2008;359:1367-80.
3. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer 1953;6:963-8.
4. Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. Cancer Prev Res (Phila Pa) 2008;1:396-403.
5. Wistuba, II, Gazdar AF. Lung cancer preneoplasia. Annu Rev Pathol 2006;1:331-48.
6. Powell CA, Klares S, O'Connor G, Brody JS. Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer. Clin Cancer Res 1999;5:2025-34.
7. Spira A, Beane J, Shah V, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. Proc Natl Acad Sci U S A 2004;101:10143-8.
8. Sridhar S, Schembri F, Zeskind J, et al. Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. BMC Genomics 2008;9:259.
9. Spira A, Beane JE, Shah V, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nat Med 2007;13:361-6.

10. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 2003;4:249-64.
11. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 2004;20:307-15.
12. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.
13. Dai M, Wang P, Boyd AD, et al. Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. Nucleic Acids Res 2005;33:e175.
14. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.
15. Tao Q, Fujimoto J, Men T, et al. Identification of the retinoic acid-inducible Gprc5a as a new lung tumor suppressor gene. J Natl Cancer Inst 2007;99:1668-82.
16. Deng J, Fujimoto J, Ye XF, et al. Knockout of the tumor suppressor gene Gprc5a in mice leads to NF-kappaB activation in airway epithelium and promotes lung inflammation and tumorigenesis. Cancer Prev Res (Phila) 2010;3:424-37.
17. Punturieri A, Szabo E, Croxton TL, Shapiro SD, Dubinett SM. Lung cancer and chronic obstructive pulmonary disease: needs and opportunities for integrated research. J Natl Cancer Inst 2009;101:554-9.
18. Borczuk AC, Sole M, Lu P, et al. Progression of human bronchioloalveolar carcinoma to invasive adenocarcinoma is modeled in a transgenic mouse model of K-ras-induced lung cancer by loss of the TGF-beta type II receptor. Cancer research 2011;71:6665-75.
19. Malkoski SP, Haeger SM, Cleaver TG, et al. Loss of transforming growth factor beta type II receptor increases aggressive tumor behavior and reduces survival in lung adenocarcinoma and squamous cell carcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research 2012;18:2173-83.
20. Li L, Wei XH, Pan YP, et al. LAPTM4B: a novel cancer-associated gene motivates multidrug resistance through efflux and activating PI3K/AKT signaling. Oncogene 2010;29:5785-95.
21. Tsai MC, Spitale RC, Chang HY. Long intergenic noncoding RNAs: new links in cancer progression. Cancer Res 2011;71:3-7.
22. Greenbaum D, Colangelo C, Williams K, Gerstein M. Comparing protein abundance and mRNA expression levels on a genomic scale. Genome Biol 2003;4:117.
23. Taniguchi Y, Choi PJ, Li GW, et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science 2010;329:533-8.
24. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 1999;19:1720-30.
25. Kettunen E, Anttila S, Seppanen JK, et al. Differentially expressed genes in nonsmall cell lung cancer: expression profiling of cancer-related genes in squamous cell lung cancer. Cancer Genet Cytogenet 2004;149:98-106.
26. Seo JS, Ju YS, Lee WC, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. Genome Res 2012.
27. Xi L, Feber A, Gupta V, et al. Whole genome exon arrays identify differential expression of alternatively spliced, cancer-related genes in lung cancer. Nucleic Acids Res 2008;36:6535-47.
28. Kikuchi T, Hassanein M, Amann JM, et al. In-depth Proteomic Analysis of Nonsmall Cell Lung Cancer to Discover Molecular Targets and Candidate Biomarkers. Mol Cell Proteomics 2012;11:916-32.

Gene expression profiling of lung tumors and matched normal airways reveals common and disparate aberrant pathways in SCC and adenocarcinoma development and potential targets for chemoprevention in early stage NSCLC patients.

Humam Kadara[1], Junya Fujimoto[1], Suk-Young Woo[3], Melinda Garcia[1], Mohamed Kabbout[1], Annette Basey[2], Jing Wang[3], Kevin R. Coombes[3], Neda Kalhor[2], Cesar Moran[2] and Ignacio I. Wistuba[1,2].

Departments of [1]Thoracic/Head and Neck Medical Oncology, [2]Pathology and [3]Bioinformatics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

Chemoprevention of non-small cell lung cancer (NSCLC) has been unsuccessful in part due to our limited knowledge of its pathogenesis. It has been suggested that normal airway epithelia share molecular abnormalities with tumors and may serve as progenitors for lung malignancies. We sought to analyze molecular profiles of lung adenocarcinomas and squamous cell carcinomas (SCCs), two major subtypes of NSCLC, and matched normal airways to elucidate aberrant expression patterns in early phases of lung tumorigenesis. All specimens were obtained from primary early stage NSCLC consented patients who did not receive neoadjuvant therapy (n=20). We profiled RNA isolated from tumors and normal lung as well as from brushings of multiple airways adjacent to NSCLC that were histopathologically confirmed to lack neoplastic or preneoplastic cells (n=194). Expression signatures signifying genes that were significantly and concurrently differentially expressed between both tumors and airways compared to normal lung tissue (tumor-airway-normal/TAN signatures) were then derived independently for SCC and smoker adenocarcinoma cases and were comprised of 1,803 and 1,938 genes, respectively. The TAN signatures effectively clustered tumor and airways from matched normal lung samples ($p<0.001$). Further analysis showed that a subset of the genes separated SCC- and adenocarcinoma-adjacent airways. Moreover, pathways and gene-network analysis using Ingenuity pathways software highlighted similarities and differences in pathway modulation between airway epithelial fields of SCCs and adenocarcinomas. Embryonic stem cell and eicosanoid signaling pathways were most significantly modulated among those common to both TAN signatures ($p<0.001$). Importantly, retinoic acid receptor and stem cell signaling pathways mediated by *NANOG* and lineage-specific oncogene, *SOX2*, were most significantly modulated in the SCC TAN signature, whereas *NF-kB* and *PTEN* signaling pathways were most prevalent in the adenocarcinoma TAN airway signature (all $p<0.001$). Gene networks mediated by decreased expression of lineage-specific oncogene, *NKX2-1/TITF-1*, and increased expression of *TP63* and networks mediated by increased expression of the *MET* and *ERBB2* oncogenes were predominantly functionally modulated in the SCC and adenocarcinoma TAN signatures, respectively. Quantitative PCR analysis confirmed the up-regulation of the *MET* oncogene in adenocarcinomas and normal airways compared to normal lung. These findings highlight expression patterns and pathways that are deregulated differentially in the pathogenesis of lung adenocarcinomas and SCCs and therefore offer therapeutic targets to guide personalized chemoprevention in early stage NSCLC patients. Supported by DoD W81XWH-10-1-1007.

Field Cancerization in Non–Small Cell Lung Cancer
Implications in Disease Pathogenesis

Humam Kadara[1] and Ignacio I. Wistuba[1,2]

[1]Department of Thoracic/Head and Neck Medical Oncology, and [2]Department of Pathology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

Lung cancer, of which non–small cell lung cancer (NSCLC) composes the majority, is the leading cause of cancer-related deaths in the United States and worldwide. NSCLCs are tumors with complex biology that we have recently started to understand with the advent of various histological, transcriptomic, genomic, and proteomic technologies. However, the histological and molecular pathogenesis of this malignancy, in particular of adenocarcinomas, is still largely unknown. Earlier studies have highlighted a field cancerization phenomenon in which histologically normal-appearing tissue adjacent to neoplastic and pre-neoplastic lesions display molecular abnormalities, some of which are in common with those in the tumors. This review will summarize advances in understanding the field cancerization phenomenon and the potential relevance of this knowledge to gain important and novel insights into the molecular pathogenesis of NSCLC as well as to subsequent development of biomarkers for early detection of lung cancers and possibly personalized prevention.

Keywords: lung cancer; field cancerization; pathogenesis; airway epithelium

Lung cancer is the leading cause of cancer deaths in the United States and worldwide in both developing and developed regions (1). The high mortality of this disease is in part due to our lacking knowledge of the molecular mechanisms governing lung cancer pathogenesis as well as the late diagnosis of the majority of lung cancers after regional or distant spread of the malignancy (2). Non–small cell lung cancer (NSCLC) represents the majority of diagnosed lung cancers (2) and is mainly composed of squamous cell carcinomas (SCCs) and lung adenocarcinomas (2, 3). Several major differences exist between adenocarcinomas and SCCs. For example, compared with SCCs that arise from the major bronchi and are mainly centrally located, pulmonary adenocarcinomas arise from small bronchi, bronchioles, or alveolar epithelial cells and are typically peripherally located, as reviewed elsewhere (2–5). Moreover, whereas SCC pathogenesis is strongly linked to smoking, lung adenocarcinoma is the more common histological subtype in never-smoker patients (6–9). Although the sequence of lesions in the pathogenesis leading to SCCs is well described, little is known about the sequential development of adenocarcinomas. Moreover, we are still lacking in our knowledge of differential mechanisms of molecular pathogenesis among both subtypes of NSCLC.

(*Received in original form January 10, 2012; accepted in final form March 1, 2012*)

Supported by Department of Defense grant W81XWH-10-1-1007 (I.I.W.).

Correspondence and requests for reprints should be addressed to Ignacio I. Wistuba, M.D., Departments of Thoracic/Head and Neck Medical Oncology and Pathology, the University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Unit 085, Houston, TX 77030. E-mail: iiwistuba@mdanderson.org.

Proc Am Thorac Soc Vol 9, Iss. 2, pp 38–42, May 1, 2012

DOI: 10.1513/pats.201201-004MS
Internet address: www.atsjournals.org

In light of the postulated period of time and multiple stages required for the development of overt epithelial tumors, it is plausible to assume that early diagnosis of lung cancer or intraepithelial lesions coupled with effective prevention strategies will reduce the significant health burden associated with this disease (10). Despite recent encouraging findings from the National Lung Screening Trial (NLST) (11), early detection and prevention of lung cancer is challenging due to the lack of biomarkers for early diagnosis of the disease and to the presence of multiple neoplastic molecular pathways that mediate lung carcinogenesis. Earlier studies have highlighted a field cancerization phenomenon in which histologically normal-appearing tissue adjacent to neoplastic and preneoplastic lesions displays molecular abnormalities, some of which are in common with those in the tumors. It is plausible to assume that understanding early events in tumor development that commence in histologically normal epithelium would pave the way for unmet effective and personalized strategies for lung cancer prevention and treatment. This review mainly summarizes advances in understanding the field cancerization phenomenon and the potential relevance of this knowledge to gain important and novel insights into the molecular pathogenesis of NSCLC as well as to subsequent development of biomarkers for early detection of lung cancers and possibly personalized prevention.

NSCLC PRE-NEOPLASIA AND MOLECULAR PATHOGENESIS

From biological and histopathological perspectives, NSCLC is a complex malignancy that develops through multiple pre-neoplastic pathways. Lung adenocarcinoma, a major subtype of NSCLC, has been increasing in incidence globally in both smokers and nonsmokers (9), with a concurrent decrease in SCC frequency. It has been suggested that the increasing incidence of lung adenocarcinomas compared with SCCs is in part due to the change in the type of cigarettes used (lower nicotine and tar) and smoking habits and behavior (7). Anatomical differences in the location of diagnosed lung adenocarcinomas and SCCs strongly suggest that both NSCLC subtypes develop through different histopathological and molecular pathways and have different cells of origin; however, the specific respiratory epithelial cell type from which each lung cancer type develops has not been established with certainty (3). However, it is noteworthy that Clara cells and the type II pneumocytes are believed to be the progenitor cells of the peripheral airways, and peripherally arising adenocarcinomas often express markers of these cell types (12, 13).

The multistage stepwise fashion of tumor development has been demonstrated in various anatomical organs exemplified by the operational stages occurring during skin carcinogenesis (14). Carcinogenesis of the skin is initiated by a carcinogen-induced mutational event, promoted by clonal outgrowth, which may be dependent on tumor promoters, followed by progression of premalignant lesions (e.g., papillomas) and their conversion to malignant tumors (14).

and 7) are associated with premalignant progression of mouse skin papillomas (14–17). The sequence of histopathological and molecular changes in bronchial epithelia that precede the development of lung SCCs have been characterized, demonstrating that sequentially occurring allele-specific molecular changes commence in dispersed foci signifying a multistage fashion of squamous lung cancer pathogenesis (4, 18). At least in this subtype of NSCLC, it has been shown that specific genomic alterations (mainly 3p21, 3p22–24, 3p25, and 9p21) occur in histologically normal bronchial epithelia from resected specimens (18). Moreover, and notably, these alterations persisted in hyperplasias, dysplasias, carcinomas *in situ*, and tumors that exhibited different commencing genomic aberrations (18). In addition, loss of heterozygosity (LOH) in the 3p region was demonstrated in normal bronchial epithelia of cancer-free smokers, further highlighting the early role this specific genomic alteration exerts in lung cancer pathogenesis (19, 20). It is noteworthy that McCaughan and colleagues demonstrated that no low-grade lesions, but all high-grade lesions, exhibited 3q amplification targeting the sex determining region Y-box 2 (*SOX2*) lineage-specific oncogene (21). These previous findings and reports, through highlighting associations between histopathological sequences and specific molecular aberrations, pinpoint to a multistage and multistep manner of lung carcinogenesis.

On the other hand, one cannot neglect the alternative hypothesis that the sequence of genetic and epigenetic alterations is irrelevant to lung cancer pathogenesis, but rather the accumulation of molecular abnormalities beyond a certain threshold mediates development of the malignant phenotype. It has been suggested that at least two molecular pathways, the Kirsten rat sarcoma viral oncogene (*KRAS*) and epidermal growth factor receptor (*EGFR*) pathways, are involved in the development of smoker and never-smoker adenocarcinomas, respectively (2, 9). Moreover, and as reviewed by Yatabe and colleagues, atypical adenomatous hyperplasias (AAHs), which are considered to be precursor lesions for peripheral lung adenocarcinomas (3, 5) and the only sequence of morphologic change identified so far for the development of invasive lung adenocarcinomas, exhibit *KRAS* mutations more frequently than invasive adenocarcinomas (5). Conversely, our group has previously demonstrated that *EGFR* mutations commence in histologically normal bronchial epithelia adjacent to lung adenocarcinomas and precede copy number increase of the oncogene (22, 23). *EGFR* mutations also are persistent throughout the different phases of lung adenocarcinoma development (5), which harbor different genomic alterations (24). It is plausible to surmise that only after increasing our knowledge of the pre-neoplastic changes as well as the corresponding molecular abnormalities leading to the development of lung adenocarcinomas would we then be able to more confidently determine whether adenocarcinomas follow a linear progression mechanism or not (5). However, and based on the aforementioned previously reported observations by our group and others, we believe that it is not counterintuitive to speculate that development of lung malignant phenotype, including that of adenocarcinomas, is due to stepwise, sequence-specific, and multistage molecular pathogenesis as well as accumulation and combination of genetic and epigenetic abnormalities.

FIELD CANCERIZATION

Smoking-Damaged Epithelium and the Field Cancerization Phenomenon

Although the majority of lung cancer patients are current or former smokers, a relatively small proportion of these smokers mary tumor development after definitive treatment by resective surgery. There is a large body of evidence that heavy smokers and patients who have survived lung cancer compose a high-risk population that may be targeted for early detection and chemoprevention efforts (4). Although the risk of developing lung cancer decreases after smoking cessation, the risk never returns to baseline. Pre-neoplastic changes have been used as surrogate endpoints for chemopreventive studies. However, it was suggested that this "shooting in the dark" approach may explain the reasons behind the general failures of clinical chemoprevention studies (10). Therefore, novel approaches to identify the best population to be targeted for early detection and chemoprevention should be devised, and risk factors for lung cancer development or relapse need to be better defined. For these important purposes, a better understanding of the biology and molecular origins of lung cancer is warranted.

Earlier work by Danely Slaughter in patients with oral cancer and oral premalignant lesions has suggested that histologically normal-appearing tissue adjacent to neoplastic and pre-neoplastic lesions display molecular abnormalities, some of which are in common with those in the tumors (25). In 1961, a seminal report by Auerbach and colleagues suggested that cigarette smoke induces extensive histological changes in the bronchial epithelia in the lungs of smokers and that premalignant lesions are widespread and multifocal throughout the respiratory epithelium, suggestive of a field effect (26). This phenomenon, coined "field of cancerization," was later shown to be evident in various epithelial cell malignancies, including lung cancer. Some degree of inflammation and inflammatory-related damage is almost invariably present in the central and peripheral airways of smokers and may precede the development of lung cancer (4). Thus, the field of cancerization may also be explained by both the direct effect of tobacco carcinogens and the initiation of inflammatory response. In this context, different theories for the origin of the field cancerization or smoking-related field of injury have been put forward and extensively reviewed elsewhere by Steiling and coworkers (27).

Multiple altered foci of bronchial epithelium are present throughout the airway in patients with lung cancer and smokers (18, 28, 29). As mentioned before, detailed analysis of histologically normal, premalignant, and malignant epithelia from patients with lung SCC indicated that multiple, sequentially occurring allele-specific chromosomal deletions of LOH commence in clonally independent foci early in the multistage pathogenesis of SCCs (18, 28). Notably, 31% percent of histologically normal epithelium and 42% of mildly abnormal (hyperplasia/metaplasia) specimens had clones of cells with allelic loss at one or more regions examined. Moreover, these molecular aberrations were also found in carcinomas *in situ* and SCCs and at a more advanced level (18). Molecular changes involving LOH of chromosomal regions 3p (e.g., fragile histidine triad gene/*FHIT*), 9p (e.g., *CDKN2A/p16*), genomic instability (increased microsatellite repeats), and *p16* methylation have been demonstrated in histologically normal bronchial epithelia in patients with SCC and in the sequence of pathogenesis of the disease (3). Moreover, Nelson and colleagues demonstrated that *KRAS* is also mutated in histologically normal lung tissue adjacent to lung tumors (30). In addition, similar epigenetic and gene methylation patterns between tumors and adjacent histologically normal epithelia were described. Belinsky and colleagues reported aberrant promoter methylation of *p16*, which was described to be commonly methylated in lung tumors (31), in at least one bronchial epithelial site from 44% of lung cancer cases examined (32). Moreover, *p16* and death-associated protein kinase (*DAPK*) promoter

cer and persisted after smoking cessation (32).

The aforementioned molecular abnormalities were detected in histologically normal epithelia adjacent to archival surgically resected tumors from patients with primary lung cancer. LOH and microsatellite alterations in multiple foci were also detected in distal histological normal bronchial epithelia of smokers without cancer (19, 20). Moreover, and importantly, these molecular abnormalities were detected in bronchial epithelia of cancer-free former smokers and appeared to have persisted for many years after smoking cessation. In addition, LOH was detected in DNA obtained from bronchial brushings of normal and abnormal lungs from patients undergoing diagnostic bronchoscopy and was detected in cells from the ipsilateral and contralateral lung (33). Mutations in *TP53* were also described to occur in bronchial epithelia of cancer-free smokers in a widely dispersed manner (34). Similar evidence also exists for promoter methylation and epigenetic changes in smoking-damaged lung epithelium of cancer-free patients. Methylation of various genes, including retinoic acid receptor 2 β (*RAR-β2*), *H-cadherin*, adenomatous polyposis coli (*APC*), *p16*, and Ras association (RalGDS/AF-6) domain family member 1 (*RASSFF1A*) has been described in bronchial epithelial cells of heavy smokers (35). Moreover, methylation of *p16*, glutathione S-transferase pi 1 (*GSTP1*), and *DAPK* was reported to be evident in bronchial brushings of one-third of cancer-free smokers examined (36). In the study by Belinsky and colleagues, as mentioned before, methylation of *p16* was detected in epithelia of cancer-free smokers (32). A more detailed list of aberrant gene promoter methylation in patients with lung cancer and cancer-free smokers is well summarized and explained in the review by Heller and coworkers (37).

High-throughput microarray profiling was shown to be useful to study the transcriptome of lung airways. Hackett and colleagues studied the expression of 44 antioxidant-related genes using bronchial brushings from cancer-free current smokers and never-smokers and found significant up-regulation of 16 of the antioxidant genes in the airways of smokers compared with nonsmokers (38). Later, Spira and colleagues described global alterations in gene expression between normal-appearing bronchial epithelium of healthy cancer-free smokers and that of nonsmokers (39). Importantly, irreversible changes in expression in airways of former smokers after years of smoking cessation were described that were believed to underlay the increased risk former smokers exhibit for developing lung cancer (39, 40). Alterations in the expression of microRNAs were also demonstrated between large airways of current and never-smokers (41). Notably, an 80-gene signature was derived from the transcriptome of large airway epithelial cells that can distinguish smokers without overt cancer from smokers with lung cancer despite originating from normal bronchial epithelia (42). More recently, Gustafson and coworkers derived a phosphoinositide-3-kinase (*PI3K*) pathway activation signature by using recombinant adenoviruses to express the 110α subunit of *PI3K* in primary human epithelial cells (43). The *PI3K* pathway activation signature was elevated in cytologically normal bronchial airways of smokers with lung cancer and, importantly, was decreased in the airways of high-risk smokers whose dysplastic lesions regressed after treatment with the *PI3K* inhibitor myoinositol (43). Microarray and gene expression profiling methodologies were also used to demonstrate the wide anatomical spread of the lung field cancerization. Common gene expression alterations were identified in bronchial, nasal, and buccal epithelia of smokers (44), and in a separate study, the expression of 119 genes was

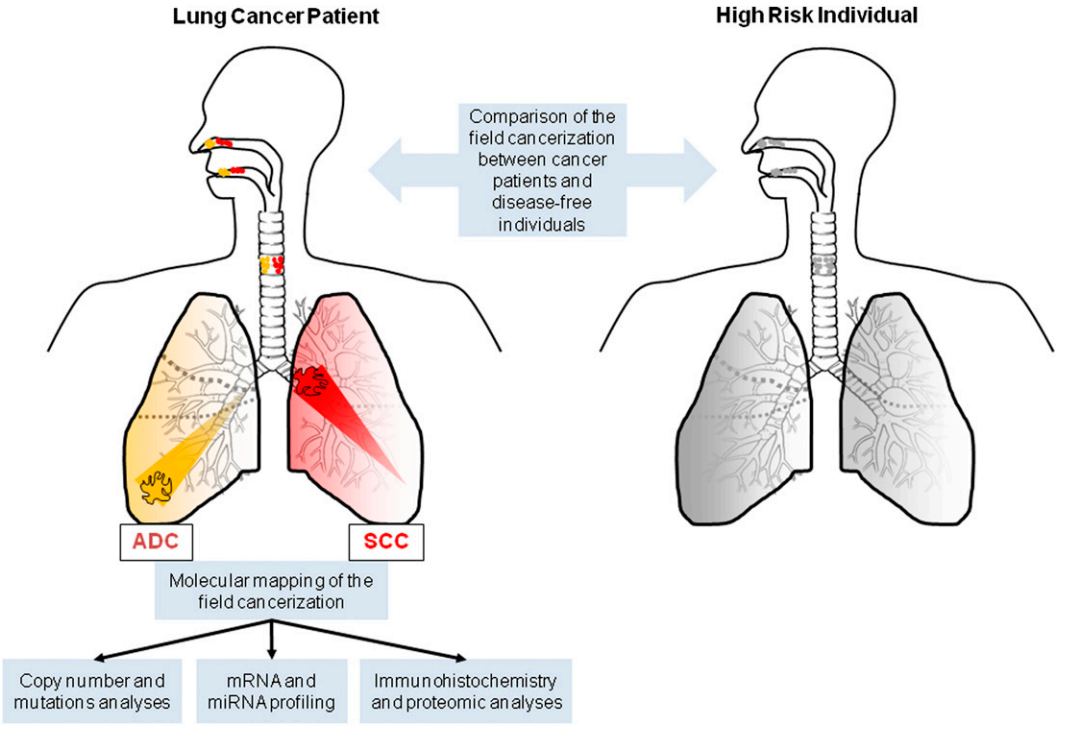


Figure 1. Molecular mapping analysis of the field cancerization in non–small cell lung cancer (NSCLC). The relevance of the lung field cancerization to the development of a particular subtype of NSCLC (i.e., adenocarcinoma compared with squamous cell carcinoma [SCC]), is still unknown, yet possible. Analyzing local and distant field of cancerization by analysis of the transcriptome of airway brushings from multiple sites independently for lung adenocarcinoma (*yellow spots*) and SCC (*red spots*) cases may shed light on events common or unique to the molecular pathogenesis of the two major subtypes of NSCLC. A "compartmental" approach coupled to a gradient or detailed molecular mapping method, which spans the tumor up to the nasal or buccal epithelium, to studying the field of cancerization may unravel biomarkers that can guide personalized prevention strategies suitable for each different NSCLC subtype. In addition, a comparison of the distant field cancerization in patients with cancer (*left*) to the expression patterns of the corresponding anatomical location in disease-free individuals (e.g., high-risk heavy smokers; *right*) would facilitate the development of efficacious markers for the detection of NSCLC. ADC = adenocarcinoma; miRNA = microRNA.

Field Cancerization Compartmentalization

In light of the prevalence of mutations in the *EGFR* oncogene in adenocarcinomas and in particular those occurring in never-smokers, Tang and colleagues investigated the presence of *EGFR* mutations in normal bronchial and bronchiolar epithelium adjacent to *EGFR* mutant tumors (22). *EGFR* mutations were detected in histologically normal peripheral epithelia in 44% of patients with lung adenocarcinoma with mutations but none in patients lacking mutations in the oncogene (22). Moreover, the same study highlighted more frequent *EGFR* mutations in normal epithelium within the tumor (43%) than in adjacent sites (24%), suggesting a localized field effect phenomenon for this abnormality in the respiratory epithelium of the lung (22). In addition, a higher frequency of mutations in cells obtained from small bronchi (35%) compared with bronchioles (18%) was detected (22). More recently, EGFR protein overexpression, similar to mutation of the gene, also exhibited a localized field effect, as it was more frequent in normal bronchial epithelia sites within tumors than in sites adjacent to and distant from tumors (23). Interestingly, *EGFR* copy number alteration was not evident in normal bronchial epithelia, which is in accordance with findings that *EGFR* copy number is a relatively late event in pathogenesis of adenocarcinomas (5, 23). These findings suggest that adenocarcinomas may be associated with a field cancerization dissimilar from that linked to SCCs.

The low frequency of molecular abnormalities detected in the centrally located bronchial respiratory epithelium in patients with peripheral lung adenocarcinomas, compared with specimens from patients with SCCs (28), suggests the presence of two compartments in the lung with different degrees of smoking-related genetic damage. Thus, smokers who develop SCCs display more smoking-related genetic damage in the respiratory epithelium of the central airway, whereas patients who develop adenocarcinomas exhibit molecular and histological damage mainly in the peripheral airways. Although some molecular changes (e.g., inflammation and signaling pathways activation) have been detected throughout the lung airway and include both compartments (central and peripheral airway), other aberrations have been more frequently altered in either central (e.g., LOH, genetic instability evidenced by microsatellite repeats) or peripheral (e.g., *EGFR* mutations as mentioned above) airways. These interesting observations indicate a possible compartmentalization of field cancerization and its dissimilarity between adenocarcinomas and SCCs, which may well reflect the differential mechanisms of pathogenesis of both NSCLC subtypes.

FUTURE DIRECTIONS AND CONCLUSION

Applying the same advanced high-throughput methodologies currently used in studying established tumors for the genetic analysis of lung NSCLC pre-neoplasia and histologically normal adjacent regions is expected to expand our understanding of the biology of this prevalent disease. Next-generation sequencing technology, through whole-genome, whole-exome, and whole-transcriptome approaches, holds great promise for providing invaluable insights into NSCLC biology, diagnosis, prevention, and therapy (46). An important step in this direction was a recent study in which RNA of bronchial airway epithelial cell brushings from healthy never-smokers and smokers with and without lung cancer was analyzed by RNA sequencing (47) and provided additional insight besides that provided when using microarray technology.

and perpetuate differential effects on the airway epithelia (4). Changes in expression in the lung field of injury have been shown to be similar in the large and small airways, and it is unknown whether they are associated with the development of the particular subtype of NSCLC. Addressing this question may be highly important, because both NSCLC subtypes display different genomic features (2) and, therefore, are clinically managed by significantly dissimilar treatment strategies, let alone differences among various subtypes of lung adenocarcinomas. Moreover, revisiting the field cancerization effect using a compartmental coupled with a gradient or detailed molecular mapping approach in patients with cancer and disease-free individuals (Figure 1) will shed light on events in the early pathogenesis of lung adenocarcinomas and SCCs and unravel biomarkers that can guide targeted and personalized chemoprevention strategies suitable for each different NSCLC subtype as well as detection efforts, in particular using less invasive sites.

Despite numerous efforts that have centered on increasing our understanding of the biology of lung cancer, this malignancy still composes the biggest share of cancer-related deaths in the United States and worldwide. Compared with advances in targeted and personalized therapy of NSCLC, little progress has been made in the tailored prevention of this fatal malignancy. This may change with the recent encouraging and significant findings of the National Lung Screening Trial (11). Various molecular markers and expression classifiers previously described in the lung airways and in less-invasive sites of field cancerization (e.g., nasal epithelium) can aid in selecting high-risk individuals best suited for computed tomography screening, for example. A comprehensive analysis of early molecular events in NSCLC pathogenesis will undoubtedly unravel biomarkers that can guide future chemoprevention strategies.

Author disclosures are available with the text of this article at www.atsjournals.org.

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;61:69–90.
2. Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med* 2008;359:1367–1380.
3. Wistuba II, Gazdar AF. Lung cancer preneoplasia. *Annu Rev Pathol* 2006;1:331–348.
4. Wistuba II. Genetics of preneoplasia: lessons from lung cancer. *Curr Mol Med* 2007;7:3–14.
5. Yatabe Y, Borczuk AC, Powell CA. Do all lung adenocarcinomas follow a stepwise progression? *Lung Cancer* 2011;74:7–11.
6. Kabat GC, Wynder EL. Lung cancer in nonsmokers. *Cancer* 1984;53: 1214–1221.
7. Khuder SA. Effect of cigarette smoking on major histological types of lung cancer: a meta-analysis. *Lung Cancer* 2001;31:139–148.
8. Rudin CM, Avila-Tang E, Harris CC, Herman JG, Hirsch FR, Pao W, Schwartz AG, Vahakangas KH, Samet JM. Lung cancer in never smokers: molecular profiles and therapeutic implications. *Clin Cancer Res* 2009;15:5646–5661.
9. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers–a different disease. *Nat Rev Cancer* 2007;7:778–790.
10. Gold KA, Kim ES, Lee JJ, Wistuba II, Farhangfar CJ, Hong WK. The battle to personalize lung cancer prevention through reverse migration. *Cancer Prev Res (Phila)* 2011;4:962–972.
11. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
12. Yatabe Y. EGFR mutations and the terminal respiratory unit. *Cancer Metastasis Rev* 2010;29:23–36.
13. Yatabe Y, Mitsudomi T, Takahashi T. TTF-1 expression in pulmonary adenocarcinomas. *Am J Surg Pathol* 2002;26:767–773.

Memorial Award Lecture. *Cancer Res* 1994;54:1178–1189.
15. Aldaz CM, Conti CJ, Larcher F, Trono D, Roop DR, Chesner J, Whitehead T, Slaga TJ. Sequential development of aneuploidy, keratin modifications, and gamma-glutamyltransferase expression in mouse skin papillomas. *Cancer Res* 1988;48:3253–3257.
16. Aldaz CM, Trono D, Larcher F, Slaga TJ, Conti CJ. Sequential trisomization of chromosomes 6 and 7 in mouse skin premalignant lesions. *Mol Carcinog* 1989;2:22–26.
17. Bianchi AB, Aldaz CM, Conti CJ. Nonrandom duplication of the chromosome bearing a mutated Ha-ras-1 allele in mouse skin tumors. *Proc Natl Acad Sci USA* 1990;87:6902–6906.
18. Wistuba II, Behrens C, Milchgrub S, Bryant D, Hung J, Minna JD, Gazdar AF. Sequential molecular abnormalities are involved in the multistage development of squamous cell lung carcinoma. *Oncogene* 1999;18:643–650.
19. Mao L, Lee JS, Kurie JM, Fan YH, Lippman SM, Lee JJ, Ro JY, Broxson A, Yu R, Morice RC, *et al*. Clonal genetic alterations in the lungs of current and former smokers. *J Natl Cancer Inst* 1997;89:857–862.
20. Wistuba II, Lam S, Behrens C, Virmani AK, Fong KM, LeRiche J, Samet JM, Srivastava S, Minna JD, Gazdar AF. Molecular damage in the bronchial epithelium of current and former smokers. *J Natl Cancer Inst* 1997;89:1366–1373.
21. McCaughan F, Pole JC, Bankier AT, Konfortov BA, Carroll B, Falzon M, Rabbitts TH, George PJ, Dear PH, Rabbitts PH. Progressive 3q amplification consistently targets SOX2 in preinvasive squamous lung cancer. *Am J Respir Crit Care Med* 2010;182:83–91.
22. Tang X, Shigematsu H, Bekele BN, Roth JA, Minna JD, Hong WK, Gazdar AF, Wistuba II. EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. *Cancer Res* 2005;65:7568–7572.
23. Tang X, Varella-Garcia M, Xavier AC, Massarelli E, Ozburn N, Moran C, Wistuba II. Epidermal growth factor receptor abnormalities in the pathogenesis and progression of lung adenocarcinomas. *Cancer Prev Res (Phila)* 2008;1:192–200.
24. Noguchi M. Stepwise progression of pulmonary adenocarcinoma–clinical and molecular implications. *Cancer Metastasis Rev* 2010;29:15–21.
25. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. *Cancer* 1953;6:963–968.
26. Auerbach O, Stout AP, Hammond EC, Garfinkel L. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. *N Engl J Med* 1961;265:253–267.
27. Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. *Cancer Prev Res (Phila)* 2008;1:396–403.
28. Wistuba II, Behrens C, Virmani AK, Mele G, Milchgrub S, Girard L, Fondon JW III, Garner HR, McKay B, Latif F, *et al*. High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. *Cancer Res* 2000;60:1949–1960.
29. Wistuba II, Berry J, Behrens C, Maitra A, Shivapurkar N, Milchgrub S, Mackay B, Minna JD, Gazdar AF. Molecular changes in the bronchial epithelium of patients with small cell lung cancer. *Clin Cancer Res* 2000;6:2604–2610.
30. Nelson MA, Wymer J, Clements N Jr. Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers. *Cancer Lett* 1996; 103:115–121.
31. Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, Baylin SB, Herman JG. Aberrant methylation of p16
32. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, Grimes MJ, Harms HJ, Tellez CS, Smith TM, *et al*. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res* 2002;62:2370–2377.
33. Powell CA, Klares S, O'Connor G, Brody JS. Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer. *Clin Cancer Res* 1999;5:2025–2034.
34. Franklin WA, Gazdar AF, Haney J, Wistuba II, La Rosa FG, Kennedy T, Ritchey DM, Miller YE. Widely dispersed p53 mutation in respiratory epithelium: a novel mechanism for field carcinogenesis. *J Clin Invest* 1997;100:2133–2137.
35. Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, Seidl S, Minna JD, Gazdar AF. Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. *Int J Cancer* 2003;107:612–616.
36. Soria JC, Rodriguez M, Liu DD, Lee JJ, Hong WK, Mao L. Aberrant promoter methylation of multiple genes in bronchial brush samples from former cigarette smokers. *Cancer Res* 2002;62:351–355.
37. Heller G, Zielinski CC, Zochbauer-Muller S. Lung cancer: from single-gene methylation to methylome profiling. *Cancer Metastasis Rev* 2010;29:95–107.
38. Hackett NR, Heguy A, Harvey BG, O'Connor TP, Luettich K, Flieder DB, Kaplan R, Crystal RG. Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers. *Am J Respir Cell Mol Biol* 2003;29:331–343.
39. Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, Palma J, Brody JS. Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci USA* 2004;101:10143–10148.
40. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol* 2007;8:R201.
41. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, Lu J, Liu G, Bowers J, Vaziri C, *et al*. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proc Natl Acad Sci USA* 2009;106:2319–2324.
42. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, Gilman S, Dumas YM, Calner P, Sebastiani P, *et al*. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* 2007;13:361–366.
43. Gustafson AM, Soldi R, Anderlind C, Scholand MB, Qian J, Zhang X, Cooper K, Walker D, McWilliams A, Liu G, *et al*. Airway PI3K pathway activation is an early and reversible event in lung cancer development. *Sci Transl Med* 2010;2:26ra25.
44. Sridhar S, Schembri F, Zeskind J, Shah V, Gustafson AM, Steiling K, Liu G, Dumas YM, Zhang X, Brody JS, *et al*. Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* 2008;9:259.
45. Zhang X, Sebastiani P, Liu G, Schembri F, Dumas YM, Langer EM, Alekseyev Y, O'Connor GT, Brooks DR, Lenburg ME, *et al*. Similarities and differences between smoking-related gene expression in nasal and bronchial epithelium. *Physiol Genomics* 2010;41:1–8.
46. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. *Nat Rev Genet* 2010; 11:685–696.
47. Beane J, Vick J, Schembri F, Anderlind C, Gower A, Campbell J, Luo L, Zhang XH, Xiao J, Alekseyev YO, *et al*. Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-seq. *Cancer Prev Res (Phila)* 2011;4:803–817.

G-Protein Coupled Receptor Family C, Group 5, Member A (*GPRC5A*) Expression Is Decreased in the Adjacent Field and Normal Bronchial Epithelia of Patients with Chronic Obstructive Pulmonary Disease and Non–Small-Cell Lung Cancer

Junya Fujimoto, MD, PhD, Humam Kadara, PhD,* Melinda M. Garcia, MS,* Mohamed Kabbout, PhD,* Carmen Behrens, MD,* Diane D. Liu, MS,† J. Jack Lee, DDS, PhD,† Luisa M. Solis, MD,‡ Edward S. Kim, MD,* Neda Kalhor, MD,‡ Cesar Moran, MD,‡ Amir Sharafkhaneh, MD, PhD,§ Reuben Lotan, PhD,*# and Ignacio I. Wistuba MD*‡*

Introduction: Understanding oncogenes and tumor suppressor genes expression patterns is essential for characterizing lung cancer pathogenesis. We have previously demonstrated that m*Gprc5a*/h*GPRC5A* is a lung-specific tumor suppressor evidenced by inflammation-mediated tumorigenesis in *Gprc5a*-knockout mice. The implication of *GPRC5A* in human lung cancer pathogenesis, including that associated with inflammatory chronic obstructive pulmonary disease (COPD), a risk factor for the malignancy, remains elusive.

Methods: We sought to examine *GPRC5A* immunohistochemical expression in histologically normal bronchial epithelia (NBE) from lung disease-free never- and ever-smokers (n = 13 and n = 18, respectively), from COPD patients with (n = 26) and without cancer (n = 24) and in non-small cell lung cancers (NSCLCs) (n = 474). Quantitative assessment of GPRC5A transcript expression in airways (n = 6), adjacent NBEs (n = 29) and corresponding tumors (n = 6) from 6 NSCLC patients was also performed.

Results: GPRC5A immunohistochemical expression was significantly lower in tumors compared to uninvolved NBE ($p < 0.0001$) and was positively associated with adenocarcinoma histology ($p < 0.001$). GPRC5A airway expression was highest in lung disease-free NBE, decreased and intermediate in NBE of cancer-free COPD patients ($p = 0.004$) and further attenuated and lowest in epithelia of COPD patients with adenocarcinoma and SCC ($p < 0.0001$). Furthermore, *GPRC5A* mRNA was significantly decreased in NSCLCs and corresponding NBE compared to uninvolved normal lung ($p = 0.03$).

Conclusions: Our findings highlight decreased *GPRC5A* expression in the field cancerization of NSCLC, including that associated with lung inflammation. Assessment of the use of *GPRC5A* expression as a risk factor for NSCLC development in COPD patients is warranted.

KEY WORDS: Field cancerization, Chronic obstructive pulmonary disease, Non–small-cell lung cancer, g-protein coupled receptor family C, group 5, member A, gene expression.

(*J Thorac Oncol.* 2012;00: 00-00)

Lung cancer, the majority of which is non–small cell lung cancer (NSCLC), is the leading cause of deaths in the United States and worldwide.[1] The high mortality associated with lung cancer is in part because of late diagnosis after regional or distant spread of the disease.[2] Improved clinical management of NSCLC is tightly linked to the identification of new effective early biomarkers that can spear novel strategies for early detection, prevention, and treatment.[2,3]

The majority (85%) of diagnosed NSCLC cases are attributable to cigarette smoking.[4,5] Auerbach et al. [6] earlier showed that tobacco carcinogen exposure causes multifocal and clonal histopathological changes in the airway epithelia of smokers, suggestive of a field cancerization in the lung. In addition, cigarette smoking perpetuates inflammation throughout the smoking-exposed airway epithelia of heavy smokers,[7] which was suggested, to lead to onset of lung cancer development.[8] Chronic obstructive pulmonary disease (COPD) of the lung is an inflammatory condition that is, like lung cancer, causally linked to cigarette smoking[9,10] and is a major cause of mortality in the United States.[11] Moreover, preinvasive lung cancer lesions are common (approximately 50%) in airways of COPD patients.[12] Importantly, although phenotypically healthy smokers comprise a significant population at risk for

Departments of *Thoracic/Head and Neck Medical Oncology, †Biostatistics and Applied Mathematics, ‡Pathology, University of Texas M.D. Anderson Cancer Center, Houston, Texas; and §Department of Medicine, Section of Pulmonary, Critical Care and Sleep Medicine, Baylor College of Medicine, Houston, Texas.

#Deceased.

J. Fujimoto and H. Kadara contributed equally to this work and should be considered first authors.

Disclosure: The authors declare no conflict of interest.

Address for correspondence: Junya Fujimoto, MD, PhD, Department of Thoracic/Head and Neck Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX. E-mail: jfujimot@mdanderson.org

ISSN: 1556-0864/12/XXXX-00

lung cancer,[5,7] COPD increases lung cancer risk by 4.5-fold, more than either age or quantity of smoking.[11,13,14] In addition, 33% of patients with mild-to-moderate COPD eventually die because of lung cancer burden.[11] Although smoking is the main cause of both diseases, only 10% to 15% of smokers develop COPD and/or lung cancer, suggesting that other factors, including genetic variation and expression, may differ across individuals in response to cigarette smoke.[4,5,9,10] It is plausible to assume that understanding common expression patterns between both diseases for subsequent identification of biological markers that explain risk of disease onset and/or progression will favorably promote the clinical management of COPD-associated lung cancer.[11]

Our group has previously demonstrated that loss of G-protein coupled receptor, family C, group 5, member A (m*Gprc5a*/h*GPRC5A*), exemplified by the *Gprc5a*-knockout mouse model, leads to chronic inflammation and spontaneous lung tumor development.[15,16] However, the expression of this tumor suppressor in human lung NSCLCs, including those associated with inflammatory conditions such as COPD, is unknown. We investigated the immunohistochemical expression of *GPRC5A* in a large series of NSCLC histological tissue specimens and in normal bronchial epithelia (NBE) from lung-disease–free individuals, COPD patients and from COPD patients with lung cancer and its transcript expression in independent resected normal epithelia and corresponding tumors. Our findings reveal marked reduced GPRC5A expression in human NSCLC and in the histologically normal field cancerization associated with COPD and lung cancer. Our study pinpoints to a potential role of this tumor suppressor in the progression of COPD-associated NSCLC, which warrants further studies to assess its use as a risk marker for this disease.

MATERIALS AND METHODS

NSCLC Tissue mMcroarrays and Normal Bronchial Epithelial Specimens

Tissue microarrays (TMAs) used in this study comprised 474 surgically resected NSCLC tumor specimens (308 adenocarcinomas, 166 squamous cell carcinomas [SCCs]) collected under an Institutional Review Board protocol and archived as formalin-fixed paraffin-embedded (FFPE) specimens in the University of Texas Specialized Program of Research Excellence thoracic tissue bank at the University of Texas M.D. Anderson Cancer Center. Clinicopathological features of the NSCLC patients examined are summarized in Table 1. TMAs were prepared with a manual tissue arrayer (Advanced Tissue Arrayer ATA100, Chemicon International, Temecula, CA) using 1-mm diameter cores in triplicate for tumors, as described previously.[17] Histological sections 4 μm in thickness were then prepared for subsequent immunohistochemistry analysis. FFPE specimens of NBE from 50 patients with COPD with forced expiratory volume in 1 second/forced vital capacity ratio of 70% or less, a smoking history of 30 or more pack-years, and collected at Baylor college of Medicine (Houston, TX) were included in the study, including 24 cancer-free cases and 26 cases with NSCLC. In addition, NBE specimens from never smokers (n = 13) and ever smokers (n = 18) without lung cancer and with no or mild airway obstruction status were analyzed.

TABLE 1. Clinicopathological Features of NSCLC Patients in Tissue Microarray Sets Used in the Study

Covariate	Levels	*n* (%)
Sex	Female	240 (50.6)
	Male	234 (49.4)
Histology	Adenocarcinoma	308 (65.0)
	Squamous cell carcinoma	166 (35.0)
Stage	I	302 (63.7)
	II	86 (18.1)
	III or IV	86 (18.1)
Grade*	Well	36 (10.4)
	Moderate	193 (55.9)
	Poor	116 (33.6)
Smoking status	Current	178 (37.6)
	Former	229 (48.3)
	Never	67 (14.1)
Tobacco history	No	67 (14.1)
	Yes	407 (85.9)

*Information on differentiation grade was available in a subset of NSCLCs analyzed.
NSCLC, non–small-cell lung cancer.

Immunohistochemical Analysis

Polyclonal antirabbit antibodies raised against human GPRC5A were generated by peptide synthesis and site-directed carrier conjugation using keyhole limpet hemocyanin as a custom service by Zymed Laboratories Inc. (South San Francisco, CA). The synthetic peptide (cysteine)-PSPYKDYEVKKEGS-COOH, corresponding to amino acids 344–357 in the human GPRC5A C-terminus, was covalently linked to keyhole limpet hemocyanin via an added cysteine residue, and polyclonal antipeptide antibodies were prepared in rabbits. Sera were confirmed to contain high-titer antibodies against the specific peptide using enzyme-linked immunosorbent assays. TMAs and the histologic sections (4 μm thick) of surgical resected specimens were deparaffinized and hydrated, and antigen retrieval was performed using a decloaker with Dako target retrieval system at a pH of 6.0 (Dako North America, Inc., Carpinteria, CA). Intrinsic peroxidase activity was blocked by 3% methanol and hydrogen peroxide for 12 minutes and serum-free protein block (Dako) was used for 30 minutes for blocking nonspecific antibody binding. Slides were then incubated with the antibodies against human GPRC5A (1:200 dilution) at 4°C overnight. After three washes in Tris-buffered saline, slides were incubated for 30 minutes with Dako Envision + Dual Link at room temperature. After three additional washes, slides were incubated with Dako chromogen substrate for 5 minutes and were counterstained with hemotoxylin for another 5 minutes. FFPE whole-section specimens, except for the omission of the primary antibodies, were used as negative controls. The intensity and extent of cytoplasmic and nuclear GPRC5A immunostaining were evaluated using a light microscope (magnification, × 20)

independently by two pathologists (JF and IIW). GPRC5A immunoreactivity was mainly cytoplasmic, which was quantified using a four-value intensity score (0, none; 1, weak; 2, moderate; and 3, strong) and the percentage (0–100%) of the extent of reactivity. A final cytoplasmic expression score was obtained by multiplying the intensity and reactivity extension values (range, 0–300).

GPRC5A Transcript Expression in Resected NSCLC Specimens and Adjacent Airway Epithelia

GPRC5A mRNA expression was assessed in ever-smoker NSCLCs and their uninvolved normal lung parenchyma tissues as well as in NBE collected by brushing of multiple corresponding airways from the same lobectomy- and pneumonectomy-resected specimens. Tumor tissues from six ever-smoker NSCLCs and normal lung specimens ($n = 6$) as well as NBE by brushings ($n = 29$) from the same patients were obtained under an Institutional Review Board approved protocol in which specimens were collected from patients who had signed letters of consent. NBE were collected from multiple adjacent to tumor airways by brushing each site independently using Cytosoft cytology brushes (Cardinal Health, Dublin, OH). Parallel brushes were used for touch-prep for cytological assessment by pan-cytokeratin staining, which revealed epithelial content more than 90%. Normal histology was determined by hematoxylin and eosin staining. Brushes were immediately placed in Qiazol lysis reagent (Qiagen, Valencia, CA) and stored in −80°C until further processing. NSCLC and normal lung specimens were shaved for histological assessment of percentage of tumor content and malignant cells and for corresponding RNA isolation. Total RNA from all samples was purified using the miRNeasy kit (Qiagen) according to the manufacturer's instructions. RNA quantity was determined using the NanoDrop spectrophotometer, (Thermo Scientific, Wilmington, DE), and quality was determined by analysis of RNA integrity with Agilent Bioanalyzer 2000 (Agilent Technologies, Santa Clara, CA).

Quantitative-Real Time Polymerase Chain Reaction

Total RNA (150 ng) was reverse-transcribed using the high-capacity RNA-to-cDNA kit (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. Quantitative real-time polymerase chain reaction (PCR) was performed using TaqMan® (Applied Biosystems, Foster City, CA) gene expression assays for *GPRC5A* (Hs00173681_m1) and beta-actin (*ACTB*) (Hs99999903_m1) primers (Life Technologies) on a 7900HT Fast-Real-Time PCR System (Life Technologies) according to the manufacturer's instructions. All samples were done in triplicates and normalized to *ACTB*. Relative quantification was calculated using the comparative cycle threshold method as previously described.[18]

Statistical Analysis

Summary statistics, including frequency tabulation, means, standard deviations, median, and range, were given to describe patient characteristics. Wilcoxon rank sum test or Kruskal–Wallis test was used to compare GPRC5A immunohistochemical expression between two levels or among more than two levels, when appropriate. General linear model was applied to test the difference of GPRC5A expression across NBE from COPD- and cancer-free never and ever smokers, cancer-free COPD patients, as well as NBE from patients with both COPD and cancer. Repeated measures analysis was used to determine significance of *GPRC5A* transcript variation across matched NSCLCs, airways, and uninvolved normal lung. All statistical tests were two-sided, and p values of 0.05 or less were considered to be statistically significant. Statistical analysis was performed with standard statistical software, including SAS Release 9.1.3 for Windows and S-Plus 8.0 for Windows.

RESULTS

GPRC5A Immunohistochemical Expression in Human NSCLC

We have previously demonstrated that m*Gprc5a*/h*GPRC5A* is a lung-specific tumor suppressor evidenced by spontaneous tumor development in mice with knockout of this gene as well as loss of the transformed phenotype in malignant NSCLC cell lines after *GPRC5A* overexpression.[16] We sought to examine the immunohistochemical expression patterns of GPRC5A tumor suppressor in NSCLC, which is unknown, and its association with clinicopathological features of the disease, including the two major histologic types examined, adenocarcinoma and SCC. The clinicopathological characteristics of the NSCLC patients from which FFPE tumor specimens were analyzed are detailed in Table 1. Representative photomicrographs of GPRC5A immunohistochemical expression in human lung adenocarcinoma (left) and SCC (right) are depicted in Figure 1*A*. GPRC5A immunoreactivity was mainly cytoplasmic. GPRC5A expression was significantly lower in the adenocarcinomas ($n = 308$; mean, 48.91 ± 36; median, 41.67; minimum, 0; maximum, 180) and SCCs ($n = 166$; mean, 23.47 ± 22.79; median, 18.33; minimum, 0; maximum, 110) examined compared with the uninvolved normal bronchial epithelia (mean, 164.52 ± 34.04; median, 160; minimum, 100; maximum, 230) from lung-disease–free smokers and never smokers ($p < 0.0001$) (Fig. 1*B*). Moreover, GPRC5A immunohistochemical expression was significantly lower in SCCs compared with that in adenocarcinomas ($p < 0.0001$) (Fig. 1*B*). We then correlated GPRC5A with other clinicopathological features for all NSCLCs. Notably, GPRC5A expression was significantly increased in NSCLC tumor specimens from never smokers ($n = 67$) (mean, 53.94 ± 39.94; median, 40; minimum, 0; maximum, 163.33) compared with ever smokers ($n = 407$) (mean, 37.71 ± 32.65; median, 30; minimum, 0; maximum, 180) ($p = 0.001$) (Fig. 1*C*). GPRC5A expression was also significantly different among never, former and current smoker NSCLCs ($p = 0.003$) (Fig. 1*D*). In addition, GPRC5A expression was significantly positively associated with well-differentiated tumor grade in a subset of NSCLC tumors ($n = 346$) examined with available grade information ($p = 0.004$) (Fig. 1*E*). It is important to note that when we examined each histology type separately, there were no statistically significant

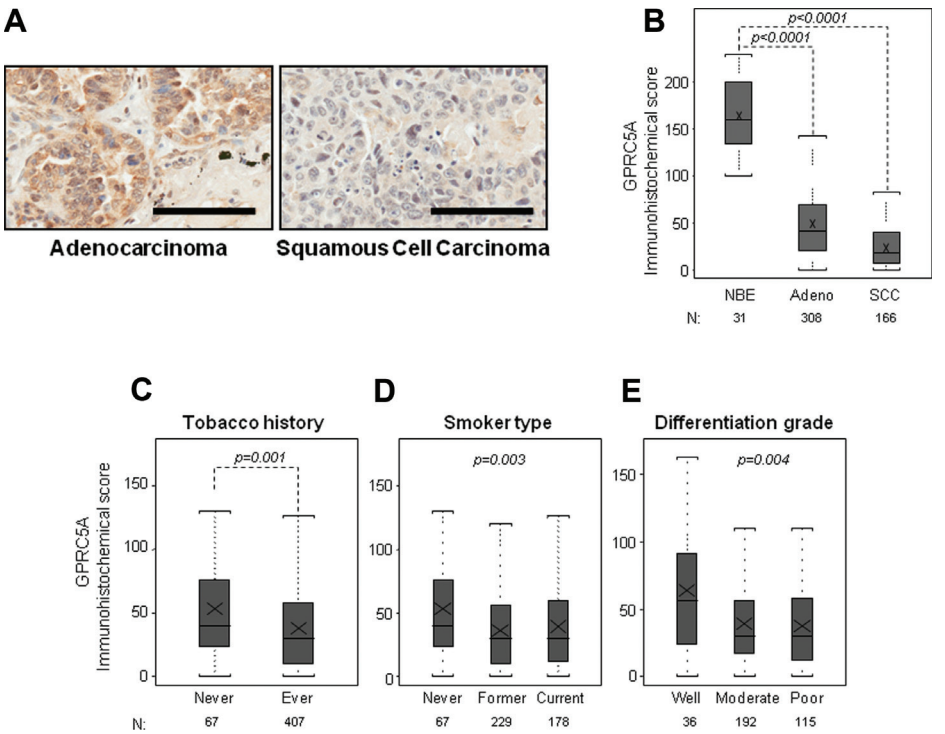


FIGURE 1. GPRC5A immunohistochemical expression in NSCLC histological tissue specimens. *A*, Representative photomicrographs of GPRC5A immunohistochemical expression in lung adenocarcinoma (left) and SCC (right) observed under a light microscope at a magnification of 20 ×. Bars, 100 μm. *B*, Box-plots depicting differences in GPRC5A immunohistochemical expression among NBE, Adeno, and SCCs. *C*, Box-plots depicting GPRC5A expression based on tobacco history (never versus ever smokers), (*D*), smoking status (never, former and current), and (*E*), differentiation grade (well, moderate, and poor). *p* values based on the Wilcoxon rank sum test and Kruskal–Wallis test. GPRC5A, G-protein coupled receptor family C, group 5, member A; NSCLC, non–small-cell lung cancer; SCC, squamous cell carcinomas; NBE, normal bronchial epithelia; Adeno, adenocarcinomas.

correlations between GPRC5A expression and patient's smoking history and status, and tumor-differentiation grade.

GPRC5A Immunohistochemical Expression in NBE from COPD Patients

We have previously shown that loss of m*Gprc5a*/h*GPRC5A* tumor suppressor leads to chronic-inflammation–mediated prosurvival signaling and transformation of nonmalignant lung epithelial cells.[15] The relevance of GPRC5A expression to human NSCLC pathogenesis, including that associated with COPD, a chronic inflammatory condition and risk factor for lung cancer, remains elusive. In light of the strong association between inflammation, COPD and NSCLC,[10,11] we sought to examine the expression of GPRC5A NBE in COPD-associated adenocarcinoma and SCC. GPRC5A expression was assessed in histological tissue specimens of NSCLC and NBE, including those from lung cancer-free COPD and from NSCLC patients. Representative photomicrographs of GPRC5A immunohistochemical staining in the different NBE are depicted in Figure 2*A*. GPRC5A immunoreactivity seemed to be highest in NBE from lung-disease–free never and ever smokers, successively significantly decreased in NBE of COPD patients who are cancer-free, and lowest in NBE of patients with both COPD and adenocarcinoma or SCC (Fig. 2*B*). A general linear model demonstrated a significant gradual decrease of GPRC5A expression from NBE of disease-free never smokers to NBE from patients with COPD and adenocarcinoma or SCC ($p < 0.0001$) (Fig. 2*B*). The mean (177.31 ± 30.66) and median (195; minimum, 130; maximum, 220) of GPRC5A expression in NBE of lung-disease–free never smokers were higher than that of smokers (mean, 155.28 ± 34.15, median, 147.5; minimum, 100; maximum, 230) although the differences were not statistically significant (Fig. 2*B*). Moreover, GPRC5A expression was significantly higher in NBE of lung-disease–free smokers relative to NBE in cancer-free COPD patients (mean, 85.48 ± 53.55, median, 72.88; minimum, 20; maximum, 211) ($p = 0.007$), which in turn was significantly higher compared with the expression in NBE from COPD patients with adenocarcinoma (mean, 48.15 ± 52.84, median, 35; minimum, 0; maximum, 170) ($p < 0.0001$) or SCC (mean, 32.24 ± 34.9, median, 12.5; minimum, 0; maximum, 101.69) ($p < 0.0001$) (Fig. 2*B*). Furthermore, although GPRC5A expression was lower in NBE of COPD patients with SCC compared with those with COPD and adenocarcinoma, the difference was not statistically significant (Fig. 2*B*). These findings suggest that reduced expression of the GPRC5A tumor suppressor may be implicated in the pathogenesis of NSCLC associated with inflammatory COPD.

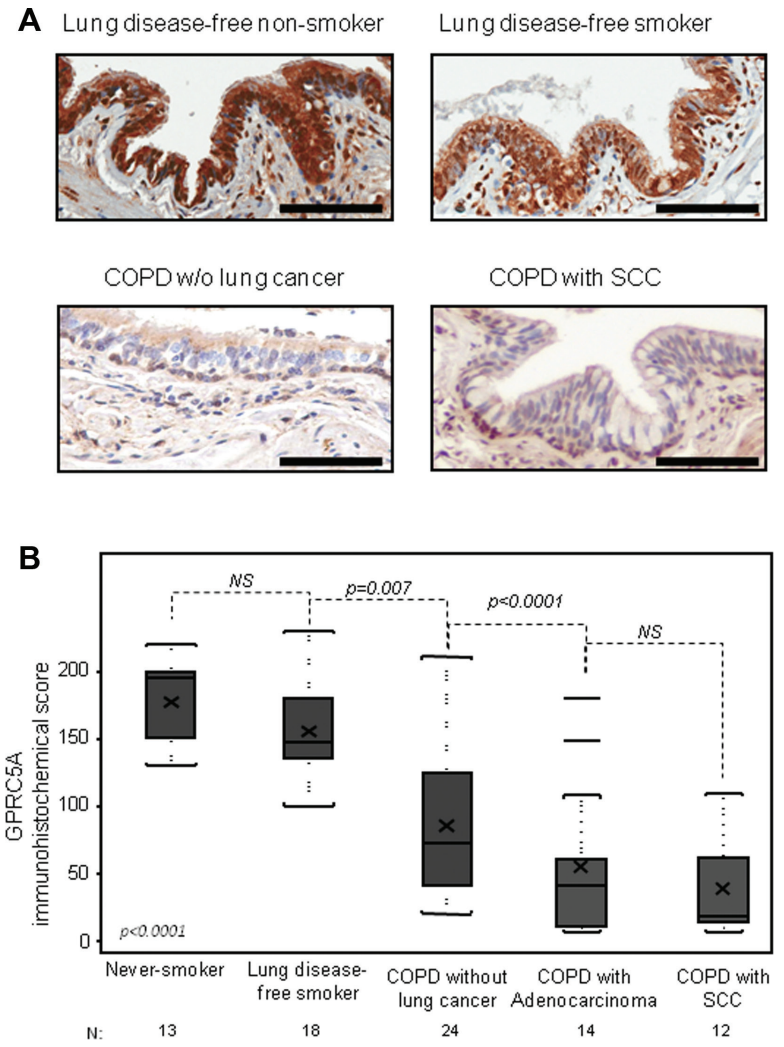


FIGURE 2. Reduced GPRC5A expression in the sequential pathogenesis of COPD-associated NSCLC. *A*, Representative photomicrographs of GPRC5A immunohistochemical expression in NBE from disease-free never and ever smokers (upper left and upper right, respectively) and in NBE from cancer-free COPD patients (bottom left) and NSCLC patients (bottom right) with COPD observed under a light microscope at a magnification of × 20. Bars, 100 μm. *B*, Box-plots depicting quantification of GPRC5A immunohistochemical expression represented in (*A*). Main *p* value signifies statistical significance of reduced GPRC5A expression in the sequential pathogenesis of COPD-associated NSCLC and among the five groups by the general linear model and Kruskal–Wallis test. *p* values signifying the indicated pair wise comparisons among the groups were obtained by the Wilcoxon rank sum test. GPRC5A, G-protein coupled receptor family C, group 5, member A; NBE, normal bronchial epithelia; NSCLC, non–small-cell lung cancer; COPD, chronic obstructive pulmonary disease; SCC, squamous cell carcinomas; NS, not significant.

We then examined GPRC5A expression in epithelial cells from different compartments of the lung (bronchial, bronchiolar and alveolar epithelia and cells). GPRC5A expression was statistically significantly lower in bronchi and bronchioles from cancer-free COPD patients ($p < 0.005$) and patients with both COPD and SCCs or adenocarcinomas ($p < 0.0001$) compared with lung-disease–free never smokers (Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/JTO/A334). In contrast, GPRC5A expression in alveolar compartment remained high (score average 172.2–261.5) in all groups of cases examined, and there were no statistically significant differences based on COPD status (Supplementary Table 1, Supplemental Digital Content 1, http://links.lww.com/JTO/A334).

GPRC5A Expression in Smoker NSCLC Field Cancerization

Our previous findings on *GPRC5A* tumor-suppressive function[15,16,19] as well as our current observation on significant and progressive reduced expression of this tumor suppressor in NBE from lung cancer-free COPD patients and patients with NSCLC, prompted us to probe whether *GPRC5A* expression exhibits a field cancerization pattern in NSCLC. We sought

to assess *GPRC5A* transcript expression in multiple bronchial sites ($n = 29$) as well as in the tumors ($n = 6$) and uninvolved normal lung parenchyma ($n = 6$) in six NSCLC cases (3 SCCs and 3 adenocarcinomas). Quantitative real-time PCR analysis demonstrated progressively reduced *GPRC5A* expression from normal uninvolved lung parenchyma (relative quantification [RQ] mean, 4.18 ± 0.62; median, 4.16; minimum, 3.51; maximum, 5.04) to histologically normal and adjacent bronchial epithelia (RQ mean, 3.12 ± 1.94; median, 2.86,; minimum, 0.59; maximum, 10.12) to NSCLC tumors (RQ mean, 1.35 ± 0.71; median 1.31; minimum 0.24; maximum 2.30), which exhibited on an average the lowest expression of the gene (Fig. 3*A* and *B*) ($p = 0.03$). These findings suggest that *GPRC5A* expression is reduced in the smoking-injured field cancerization in NSCLC highlighting a potentially strong implication of this tumor suppressor in the pathogenesis of this malignancy.

DISCUSSION

In this study, we assessed the expression of *GPRC5A*, which we had found to exhibit lung-specific tumor-suppressor properties in mice,[15,16,19] in human NSCLC histological tissue specimens, as well as in NBE from COPD and lung adenocarcinoma and SCC patients. We found that the protein product of human *GPRC5A* was lower in lung tumors compared with uninvolved histologically NBE, and was significantly associated with adenocarcinoma histology. *GPRC5A* airway expression seemed to exhibit a progressive decrease in the field cancerization of COPD-related NSCLC with highest expression in bronchial epithelia obtained from disease-free individuals, intermediate immunoreactivity in normal epithelia from cancer-free COPD patients, and lowest in normal epithelia from patients with both COPD and adenocarcinoma and SCC histologies. Furthermore, *GPRC5A* transcript expression was also lower in histologically NBE as well as in adjacent and corresponding smoker NSCLC tumors, irrespective of COPD status, compared with uninvolved normal lung suggestive of a field cancerization-mediated expression modulation of this G-protein coupled receptor. In light of the relevance of the field cancerization phenomenon to NSCLC pathogenesis,[5–7,20,21] our findings pinpoint to a potential tumor-suppressive role, similar to that established in mice, of *GPRC5A* in the sequential development of human NSCLC, in particular those associated with inflammatory chronic obstructive disease. NSCLC and COPD are both mainly attributable to cigarette smoke[9,10,13] and are leading causes of deaths in the United States and worldwide.[1,2,11] Given that COPD is a major risk factor for lung cancer and shares various pathogenic features with lung tumors,[9,10,22–25] better molecular markers are needed to identify which COPD patients will continue to develop lung malignancies.[11] Thus, the results herein raise the intriguing possibility that *GPRC5A* loss may be a useful biomarker in assessing the risk of NSCLC development in COPD patients.

COPD and infections as well as inflammatory disorders of the respiratory tract may be linked to an increased risk of lung cancer.[26,27] We have previously demonstrated significantly increased macrophage infiltration into lungs of *Gprc5a*-knockout mice along with their direct association with adenocarcinomas, which was accompanied by higher constitutive levels of proinflammatory cytokines and chemokines and increased susceptibility to stimulation of nuclear factor-kappa B (NF-κB) activation in vivo.[15] We also showed that loss of *Gprc5a*-mediated activation of NF-κB was causally linked to macrophage recruitment and enhanced inflammation toward the creation of a tumor-promoting microenvironment.[15] It is noteworthy that lungs of patients with severe COPD exhibit macrophage/$CD8^+$ T cell and neutrophil infiltration based on histopathological studies and bronchial-biopsies.[28] Moreover, it is worthwhile to mention that we attempted to statistically assess the association between expression of GPRC5A and that of NF-κB (nuclear p65 immunoreactivity) in NSCLCs and found a significant

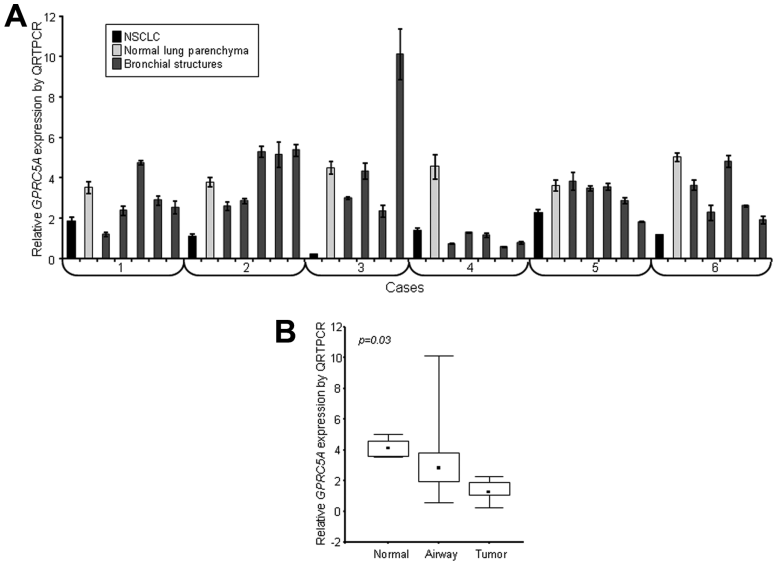


FIGURE 3. Decreased *GPRC5A* mRNA expression in histologically NBE and corresponding NSCLC tumors from resected specimens compared with matched uninvolved normal lung. Total RNA was isolated from brushings of NBE and frozen sections of matched NSCLC (T) and uninvolved NL parenchyma obtained from resected lobectomy or pneumonectomy specimens performed on smoker early-stage patients. *A, GPRC5A* expression was analyzed in all samples (in triplicates) by QRTPCR and normalized against that of *ACTB* to yield a relative expression by the $2^{-\Delta\Delta CT}$ method depicted in the graphs. NSCLC tumors (squamous cell carcinoma, cases 1–3, and adenocarcinomas, cases 4–6), uninvolved NL parenchyma and bronchial structures are labeled by the indicated colors. *B, GPRC5A* transcript expression among uninvolved NL tissue, NBE, and NSCLC tumors was assessed statistically by repeated measures analysis. NBE, normal bronchial epithelia; NSCLC, non–small-cell lung cancer; NL, normal lung; CT, comparative cycle threshold method.; QRTPCR, quantitative real-time polymerase chain reaction.

inverse correlation between both proteins in particular in adenocarcinomas (data not shown). In addition, it was recently shown that the incidence of hyperplastic lesions in lungs of *Gprc5a*-knockout mice was increased after exposure to nontypeable *Haemophilus* influenzae.[29] Notably, bacterial colonization, particularly with nontypeable *Haemophilus* influenzae, has been implicated as a cause of airway inflammation in COPD besides cigarette smoke.[30] In this context, the commonalities between inflammatory-mediated and histopathological mechanisms in *Gprc5a*-knockout mice and those evident in lungs of COPD patients along with our current findings on reduced *GPRC5A* expression in normal cells from COPD patients pinpoint to a tumor-suppressive role of this gene in COPD-associated human lung tumorigenesis.

We previously cloned the retinoid-regulated m*Gprc5a*/h*GPRC5A* and found that it was preferentially expressed in fetal and adult mouse and human lung tissue compared with normal specimens from other anatomically distinct organs.[31] The functional relevance of this lung-specific expression was highlighted by our earlier study demonstrating spontaneous development of adenomas and adenocarcinomas in *Gprc5a*-nockout mice, which was not observed in wild-type littermates. In addition, spontaneous tumor development in *Gprc5a*-knockout mice was characterized by late tumor onset (12–16-month-old mice) and low multiplicity.[16] Notably, we reported that exposure to the tobacco-specific carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (nicotine-derived nitrosamine ketone) augmented tumorigenesis evidenced by 6-month earlier development of lesions, two- to fourfold increased tumor incidence and multiplicity, as well as a dramatic increase in lesion size.[19] It is worthy to note that, in this study, we found significantly reduced expression of *GPRC5A* in human smoking-injured histologically NBE adjacent to lung tumors and in the localized field cancerization of smoker NSCLC patients, compared with uninvolved normal lung. Moreover, reduced *GPRC5A* expression was concomitant among airways and corresponding NSCLC compared with matched normal lung tissue. It is plausible to assume that smoking-mediated reduced airway expression of *GPRC5A* in patients may explain our previous observations on augmented tumorigenesis in *Gprc5a*-knockout mice after exposure to tobacco carcinogens. Because COPD is causally linked to smoking,[9–11,22,23] it cannot be neglected that reduced *GPRC5A* expression in NBE from COPD patients may be because of the molecular impact of smoking on the airway transcriptome. However, it is important to mention that we found significantly decreased *GPRC5A* expression in normal epithelia from COPD patients compared with epithelia obtained from disease-free smokers. Moreover, GPRC5A expression was further decreased and lowest in normal airway epithelia from COPD patients with NSCLC compared with those from cancer-free COPD patients. Thus in this context, our findings raise the possibility that *GPRC5A* expression may be impacted by cigarette smoking and may be both a cause and consequence of increased inflammation in tumor-promoting microenvironment in the lung as well as NSCLC development, which may be better delineated in future warranted studies.

Our group has previously shown that human lung and head and neck cancer cell lines exhibit reduced *GPRC5A* expression compared with their normal counterparts.[31] It is worthwhile to mention that treatment of cancer cells with all-trans retinoic acid restored GPRC5A expression in cells devoid of the tumor suppressor in part by a retinoic acid receptor-dependent manner.[31,32] In addition and in a separate study, we have previously noted reduced retinoic acid receptor beta (RARβ) in approximately 50% of adenocarcinomas and SCCs as well as loss of RARγ and retinoid X receptorβ in a significant fraction of the tumors.[33] Moreover, RARβ expression is decreased in bronchial epithelia and preneoplastic lesions representing the sequence of lung SCC pathogenesis.[34] More recently, we demonstrated, using gene expression profiling, that loss of murine Gprc5a in lung epithelial cells reduced expression of markers of squamous differentiation concomitantly with an increase in mediators of the inflammatory process (e.g., NF-κB signaling). In light of the biological connection between retinoid signaling and GPRC5A expression as well as our current findings in this revised article on the significant reduced expression of the tumor suppressor in NBE of COPD patients including those with NSCLC, it is plausible to suggest that retinoid signaling may be aberrantly regulated in pathogenesis of COPD-associated NSCLC. It is important to note that retinoids are currently being tested for their lung regenerative properties and effects in patients with COPD.[35]

In conclusion, we report herein decreased expression of *GPRC5A* in human NSCLC tissue compared with uninvolved NBE and in the field cancerization of smoker NSCLCs pinpointing to a tumor-suppressor role of this G-protein coupled receptor in the pathogenesis of this leading cause of cancer-related deaths. Moreover, our findings highlight a progressive decrease in *GPRC5A* expression in the sequential pathogenesis of NSCLCs arising in COPD patients, warranting future studies, including the analysis of lung cancer preneoplastic lesions, to assess the potential of the utility of this gene as a biomarker for lung cancer risk in COPD patients.

ACKNOWLEDGMENTS

This study was supported in part by grants from the Samuel Waxman Cancer Research Foundation and Department of Defense (DoD) grants W81XWH-04-1-0142 and W81XWH-10-1-1007 (to Ignacio I. Wistuba, MD), and the National Institute of Health, Specialized Program of Research Excellence (SPORE) in lung cancer P50CA70907 (to Ignacio I. Wistuba), and the Cancer Center Support Grant (CCSG) (CA-16672).

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin* 2012;62:10–29.
2. Herbst RS, Heymach JV, Lippman SM. Lung cancer. *N Engl J Med* 2008;359:1367–1380.
3. Goldstraw P, Ball D, Jett JR, et al. Non-small-cell lung cancer. *Lancet* 2011;378:1727–1740.
4. Gazdar AF, Thun MJ. Lung cancer, smoke exposure, and sex. *J Clin Oncol* 2007;25:469–471.
5. Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. *Cancer Prev Res (Phila)* 2008;1:396–403.
6. Auerbach O, Stout AP, Hammond EC, Garfinkel L. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. *N Engl J Med* 1961;265:253–267.

7. Wistuba II, Gazdar AF. Lung cancer preneoplasia. *Annu Rev Pathol* 2006;1:331–348.
8. David H. Rudolf Virchow and modern aspects of tumor pathology. *Pathol Res Pract* 1988;183:356–364.
9. Fletcher C, Peto R. The natural history of chronic airflow obstruction. *Br Med J* 1977;1:1645–1648.
10. Mannino DM. COPD: epidemiology, prevalence, morbidity and mortality, and disease heterogeneity. *Chest* 2002;121(5 Suppl):121S–126S.
11. Punturieri A, Szabo E, Croxton TL, Shapiro SD, Dubinett SM. Lung cancer and chronic obstructive pulmonary disease: needs and opportunities for integrated research. *J Natl Cancer Inst* 2009;101:554–559.
12. Lam S, leRiche JC, Zheng Y, et al. Sex-related differences in bronchial epithelial changes associated with tobacco smoking. *J Natl Cancer Inst* 1999;91:691–696.
13. Mannino DM, Aguayo SM, Petty TL, Redd SC. Low lung function and incident lung cancer in the United States: data From the First National Health and Nutrition Examination Survey follow-up. *Arch Intern Med* 2003;163:1475–1480.
14. Tockman MS, Anthonisen NR, Wright EC, Donithan MG. Airways obstruction and the risk for lung cancer. *Ann Intern Med* 1987;106:512–518.
15. Deng J, Fujimoto J, Ye XF, et al. Knockout of the tumor suppressor gene Gprc5a in mice leads to NF-kappaB activation in airway epithelium and promotes lung inflammation and tumorigenesis. *Cancer Prev Res (Phila)* 2010;3:424–437.
16. Tao Q, Fujimoto J, Men T, et al. Identification of the retinoic acid-inducible Gprc5a as a new lung tumor suppressor gene. *J Natl Cancer Inst* 2007;99:1668–1682.
17. Sun M, Behrens C, Feng L, et al. HER family receptor abnormalities in lung cancer brain metastases and corresponding primary tumors. *Clin Cancer Res* 2009;15:4829–4837.
18. Tahara E, Kadara H, Lacroix L, Lotan D, Lotan R. Activation of protein kinase C by phorbol 12-myristate 13-acetate suppresses the growth of lung cancer cells through KLF6 induction. *Cancer Biol Ther* 2009;8:801–807.
19. Fujimoto J, Kadara H, Men T, van Pelt C, Lotan D, Lotan R. Comparative functional genomics analysis of NNK tobacco-carcinogen induced lung adenocarcinoma development in Gprc5a-knockout mice. *PLoS ONE* 2010;5:e11847.
20. Braakhuis BJ, Tabor MP, Kummer JA, Leemans CR, Brakenhoff RH. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. *Cancer Res* 2003;63:1727–1730.
21. Gold KA, Kim ES, Lee JJ, Wistuba II, Farhangfar CJ, Hong WK. The BATTLE to personalize lung cancer prevention through reverse migration. *Cancer Prev Res (Phila)* 2011;4:962–972.
22. Cohen BH, Diamond EL, Graves CG, et al. A common familial component in lung cancer and chronic obstructive pulmonary disease. *Lancet* 1977;2:523–526.
23. Rooney C, Sethi T. The epithelial cell and lung cancer: the link between chronic obstructive pulmonary disease and lung cancer. *Respiration* 2011;81:89–104.
24. Schottenfeld D, Beebe-Dimmer J. Chronic inflammation: a common and important factor in the pathogenesis of neoplasia. *CA Cancer J Clin* 2006;56:69–83.
25. Schwartz AG, Ruckdeschel JC. Familial lung cancer: genetic susceptibility and relationship to chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2006;173:16–22.
26. Engels EA. Inflammation in the development of lung cancer: epidemiological evidence. *Expert Rev Anticancer Ther* 2008;8:605–615.
27. Lee G, Walser TC, Dubinett SM. Chronic inflammation, chronic obstructive pulmonary disease, and lung cancer. *Curr Opin Pulm Med* 2009;15:303–307.
28. Saetta M, Di Stefano A, Turato G, et al. CD8+ T-lymphocytes in peripheral airways of smokers with chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 1998;157(3 Pt 1):822–826.
29. Barta P, Van Pelt C, Men T, Dickey BF, Lotan R, Moghaddam SJ. Enhancement of lung tumorigenesis in a Gprc5a Knockout mouse by chronic extrinsic airway inflammation. *Mol Cancer* 2012;11:4.
30. Hallström T, Singh B, Resman F, Blom AM, Mörgelin M, Riesbeck K. Haemophilus influenzae protein E binds to the extracellular matrix by concurrently interacting with laminin and vitronectin. *J Infect Dis* 2011;204:1065–1074.
31. Cheng Y, Lotan R. Molecular cloning and characterization of a novel retinoic acid-inducible gene that encodes a putative G protein-coupled receptor. *J Biol Chem* 1998;273:35008–35015.
32. Ye X, Tao Q, Wang Y, Cheng Y, Lotan R. Mechanisms underlying the induction of the putative human tumor suppressor GPRC5A by retinoic acid. *Cancer Biol Ther* 2009;8:951–962.
33. Xu XC, Sozzi G, Lee JS, et al. Suppression of retinoic acid receptor beta in non-small-cell lung cancer in vivo: implications for lung cancer development. *J Natl Cancer Inst* 1997;89:624–629.
34. Xu XC, Lee JS, Lee JJ, et al. Nuclear retinoid acid receptor beta in bronchial epithelium of smokers before and during chemoprevention. *J Natl Cancer Inst* 1999;91:1317–1321.
35. Hind M, Gilthorpe A, Stinchcombe S, Maden M. Retinoid induction of alveolar regeneration: from mice to man? *Thorax* 2009;64:451–457.

Characterizing the molecular spatial and temporal field of injury in early stage smoker non-small cell lung cancer patients after definitive surgery by expression profiling

Humam Kadara[1], Li Shen[2], Junya Fujimoto[1], Pierre Saintigny[1], Chi-Wan Chow[1], Wenhua Lang[1], Zuoming Chu[1], Melinda Garcia[1], Mohamed Kabbout[1], You-Hong Fan[1], Carmen Behrens[1], Diane A. Liu[4], Li Mao[5], J. Jack Lee[4], Kathryn A. Gold[1], Jing Wang[2], Kevin R. Coombes[2], Edward S. Kim[1‡], Waun Ki Hong[1], Ignacio I. Wistuba[1,3*].

Departments of [1]Thoracic/Head and Neck Medical Oncology, [2]Bioinformatics, [3]Pathology and [4]Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. [5]School of Dentistry, The University of Maryland, Baltimore, Baltimore, MD, USA.

*Correspondence: Ignacio I. Wistuba, M.D., Departments of Thoracic/Head and Neck Medical Oncology and Pathology, the University of Texas MD Anderson Cancer Center, TX, USA, Tel: 713-563-9184, Fax: 713-730-0309, email: iiwistuba@mdanderson.org.

‡Present address: Department of Solid Tumor Oncology and Investigational Therapeutics, Levine Cancer Institute, Carolinas Healthcare System, Charlotte, NC, USA, email: Edward.Kim@carolinashealthcare.org.

Grant support

This study was funded in part by Department of Defense (DoD) grants W81XWH-04-1-0142 (to W.K.H. and I.I.W.) and W81XWH-10-1-1007 (to H.K. and I.I.W.) and by Cancer Center Support Grant CA16672 (MD Anderson Cancer Center microarray core facility).

Keywords

Early stage NSCLC, gene expression profiling, lung airway epithelium, chemoprevention

Conflict of Interest:

No conflicts declared

Abstract

Gene expression alterations in response to cigarette smoke have been characterized in normal-appearing bronchial epithelium of healthy smokers and it has been suggested that adjacent histologically normal tissue display tumor-associated molecular abnormalities. We sought to delineate the spatial and temporal molecular lung field of injury in smoker early stage non-small cell lung cancer (NSCLC) patients (n=19) who were accrued into a surveillance clinical trial for annual follow-up and bronchoscopies within one year after definitive surgery. Bronchial brushings and biopsies were obtained from six different sites in the lung at the time of inclusion in the study and at 12, 24 and 36 months after the first time point. Affymetrix Human Gene 1.0 ST arrays were used for whole-transcript expression profiling of airways (n=391). Microarray analysis identified gene features (n=1165) that were non-uniform by site and differentially expressed between airways adjacent to tumors relative to more distant samples as well as those (n=1395) that were significantly altered with time up to three years. In addition, gene-interaction networks mediated by PI3K and ERK1/2 were modulated in adjacent compared to contralateral airways and the latter network with time. Furthermore, phosphorylated AKT and ERK1/2 immunohistochemical expression were significantly increased with time (nuclear pAKT, $p=0.03$; cytoplasmic pAKT, $p<0.0001$; pERK1/2, $p=0.02$) and elevated in adjacent compared to more distant airways (nuclear pAKT, $p=0.04$; pERK1/2, $p=0.03$). This study highlights spatial and temporal cancer-associated expression alterations in the molecular field of injury of early stage NSCLC patients after definitive surgery that warrant further validation in independent studies.

Introduction

Lung cancer, of which non-small cell lung cancer (NSCLC) comprises the majority, is the leading cause of cancer-related deaths in the US and worldwide (1, 2). The high mortality of this disease is in part due to the late diagnosis of the majority of lung cancers after regional or distant spread of the malignancy (3). Recent data from the National Lung Screening Trial (4), indicating that screening increases early detection rates, are expected to augment the number of early stage NSCLC detected warranting the need for better clinical management of this growing subpopulation. Besides adjuvant therapy, there are no effective chemoprevention strategies for early stage NSCLC patients (5) who comprise approximately 50% of all diagnosed cases and have a relatively high rate of relapse (3). Improved clinical management of early stage NSCLC is tightly linked to the identification of new effective early biomarkers that can guide potential chemoprevention strategies.

Most diagnosed NSCLC cases (85%) are attributable to cigarette smoking (6-8). Auerbach *et al* found that cigarette smoke induces extensive histological changes in the bronchial epithelia in the lungs of smokers and that premalignant lesions are widespread and multifocal throughout the respiratory epithelium, suggestive of a field effect (9). Many molecular abnormalities, such as loss of heterozygosity (LOH) (10-12), mutations in *TP53* (13), methylation of *p16* tumor suppressor, death associated protein kinase (*DAPK*) and retinoic acid receptor 2 beta (*RAR-β2*) were detected in bronchial epithelia of cancer-free former smokers (14-18) some of which persisting for many years after smoking cessation (15). More recently, global mRNA and microRNA (miRNA) expression profiles have been described in the normal-appearing bronchial epithelium of healthy smokers (19, 20) that are different from those in non-smokers. Moreover, expression and pathway signatures have been derived from normal

bronchial epithelium of smokers that exhibited diagnostic properties (21, 22). Molecular changes involving LOH of chromosomal regions 3p (*DDUT* and *FHIT* genes), 9p (*CDKN2A*), genomic instability (increased microsatellite repeats) and *p16* methylation have been demonstrated in histologically normal epithelium in squamous cell carcinoma patients and in the sequence of pathogenesis of the disease (14, 23, 24). Moreover, Nelson *et al* demonstrated that *KRAS* is also mutated in histologically normal lung tissue adjacent to lung tumors (25). Furthermore, Tang and colleagues found higher rates of *EGFR* mutations in adjacent normal bronchial epithelia (26, 27) suggestive of a potential localized field effect.

It is plausible to assume that understanding early molecular aberrations in histologically normal smoke-damaged airway epithelium of early stage patients would serve as a critical first step towards identification of biomarkers that can guide lung cancer prevention strategies. However, the global molecular airway field of injury in early stage NSCLC patients, in particular after definitive surgery, is unknown. In this study, we used transcript-level expression profiling coupled with gene-interaction network analysis and immunohistochemical analysis to characterize, in-depth, site- and time-dependent global molecular alterations in airways of smoker early stage NSCLC patients.

Materials and Methods

Patient population and airway epithelial cell collection. Early stage (I/II), current or former smoker NSCLC patients with at least a 10-pack-year smoking history and without evidence of disease after definitive surgery were recruited into the Vanguard phase II surveillance clinical trial (clinical trial number NCT00352391) within one year from time of surgery. Patients were accrued between 2004-2008 and underwent frequent testing including chest x-rays, CT scans,

laboratory work, serologies, flexible bronchoscopies and airway biopsy collections within one year from surgery (average, 6 months; range, 1-12) (first time point), and at months 12, 24, and 36 following the first time point. Bronchoscopies were performed using white light or both white light and autofluorescence modalities. Biopsies were obtained from all potential anatomical locations and time points per patient. In total there were 324 evaluable airway biopsies. Histological assessment was performed to determine whether malignant changes will occur during the time period. Once patients have completed 3 years of testing, they were followed until the study is completed. Patients were comprised of former (n=16) and current (n=3) smokers. One of the three current smoker patients quit smoking 6 months before the 24 month time point. The clinicopathological variables of patients in this study are summarized in Table 1. The study was approved by the Institutional Review Boards and all participants provided written informed consents.

Bronchial airway epithelial cells were obtained from five to six different sites (main carina, 4 airways from 4 lobes and the bronchial stump or main stem bronchus adjacent to the originally resected tumor and lobe, Figure 1) at each time point using an Olympus fiberoptic bronchoscope (Olympus America Inc., Center Valley, PA) and cytobrushes (Cellebrity Endoscopic Cytobrush, Boston Scientific, Boston, MA). Patients (n=19) with samples/specimens available for analysis that were obtained serially up to either 24 or 36 months and from at least from four different sites in the lung at each time point (n=391 airway samples) were selected for the study. Epithelial cell content was confirmed by cytokeratin staining which yielded a 90% epithelial cell content mainly comprised of ciliated cells. Brushes were immediately placed in serum-free RPMI medium on ice, vortexed gently to disperse epithelia into the media and then removed. Samples were then immediately centrifuged and cell pellets were resuspended in 1 ml

of phosphate-buffered saline (PBS). 500 µl of the sample was then again centrifuged and the pellet resuspended in 500 µl β-mercaptoethanol-containing RLT buffer (Qiagen, Valencia, CA), homogenized and stored in -80°C until further processing. Total RNA was isolated using the RNeasy Mini Kit according to the manufacturer's instructions (Qiagen).

RNA processing for microarrays. Total RNA samples were preprocessed for subsequent hybridization to expression arrays using the WT-Ovation and Encore™ Biotin Module from NuGEN Technologies Inc. (San Carlos, CA) according to the manufacturer's instructions. Briefly, the WT-Ovation Pico RNA amplification system was used to generate amplified cDNA using 5 ng of starting RNA material. After formation of double stranded cDNA, DNA was amplified using the SPIA™ amplification method, a linear isothermal DNA amplification process developed by the vendor (NuGEN Technologies). The WT-Ovation™ Exon Module (NuGEN Technologies) was then used for generation of amplified sense strand cDNA (ST-cDNA) that is suitable for subsequent array analysis with the Affymetrix Human Gene 1.0 ST array platform (Affymetrix, Santa Clara, CA). Fragmented and biotin-labeled cDNA was then generated using the Encore™ Biotin Module (NuGEN Technologies) using 5 µg of amplified cDNA. Quality and size distribution of unfragmented SPIA-amplified cDNA and subsequent fragmented labeled cDNA were assessed by loading samples on an RNA 6000 Nano LabChip (Agilent) and analysis with Agilent bioanalyzer 2000 (Agilent). No differences in quality were noted based on duration of sample storage.

Generation of microarray raw data and analysis. Fragmented and labeled cDNA (2.5 µg) were hybridized onto Human Gene 1.0 ST arrays according to the manufacturer's instructions

(Affymetrix). Hybridization cocktails containing samples, control oligonucleotide and eukaryotic hybridization controls in addition to hybridization mixes, DMSO and nuclease-free water were heat denatured at 99^{o}C for 5 minutes, cooled to 45^{o}C for 5 minutes, and finally centrifuged at maximum speed for 1 minute. After injecting 80 μL of the hybridization cocktails, arrays were incubated for 17 hours in a hybridization oven set to a temperature of 45^{o}C with 60 rpm rotation. Arrays were washed, stained and processed using Affymetrix GeneChip Fluidics Station 450 systems after which they were imaged using Affymetrix GeneChip Scanner 3000 7G for subsequent generation of raw data (*CEL files). Raw data were quantified using Robust Multichip Array (RMA) background correction, quantile normalization and RMA probe-level models (RmaPlm) summarization methods. MIAME compliant metadata, normalized expression values and 391 CEL files were submitted to the gene expression omnibus (GEO) (samples GSM992943-GSM993345, series GSE40407). After data preprocessing and normalization, a log base 2 transformation was applied. Pathways and gene-interaction network analyses were performed using the commercially available software Ingenuity Pathways analysis. All details of the microarray analysis including pairwise analysis of adjacent and contralateral airways for patient clustering are included in the Supplementary Information and in the four supplementary sweave reports accompanying the manuscript.

Immunohistochemical analysis of airway biopsies. Immunohistochemistry was done on histological sections of 4 micron formalin-fixed paraffin-embedded tissue samples prepared by a tissue arrayer as described previously (28). Immunohistochemistry analysis was performed using purified rabbit polyclonal primary antibodies raised against phospho-AKT(Threonine308) (1:200 dilution, clone C31E5E, catalog number 2965) and phospho-ERK1/2(Thr202/Tyr204) (1:400 dilution, clone D13.14.4E, catalog number 4370) (Cell Signaling Technology, Danvers, MA).

Antigen retrieval was performed using the Dako target retrieval system at a PH of 6 (Dako, Carpinteria, CA). Intrinsic peroxidase activity was blocked by 3% methanol and hydrogen peroxide for 15 min and serum-free protein block (Dako) was used for 7 min for blocking non-specific antibody binding. Slides were then incubated with the antibodies against phospho-AKT and phospho-ERK1/2 at room temperature for 90 and 120 min, respectively. After three washes in Tris-buffered saline, slides were incubated for 30 min with Dako Envision+ Dual Link at room temperature. Following three additional washes, slides were incubated with Dako chromogen substrate for 5 min and were counterstained with hemotoxylin for another 5 min. Formalin-fixed and paraffin-embedded pellets from lung cancer cell lines displaying positive phospho-AKT and phospho-ERK1/2 expression by Western blot analysis were used as a positive control, whereas samples and whole-section tissue specimens processed similarly, except for the omission of the primary antibodies were used as negative controls. The intensity and extent of cytoplasmic and nuclear phospho-AKT and phospho-ERK1/2 immunostaining was evaluated using a light microscope (magnification, x20) independently by two pathologists (J.F. and I.I.W.) who were blinded to the identity of the samples. Cytoplasmic expression was quantified using a four-value intensity score (0, none; 1, weak; 2, moderate and 3, strong) and the percentage (0-100%) of the extent of reactivity). A final cytoplasmic expression score was obtained by multiplying the intensity and reactivity extension values (range, 0-300). Nuclear expression score was quantified using the percentage of extent of reactivity (range, 0-100).

Summary statistics, including frequency tabulation, means, standard deviations, median and range were given to describe subject characteristics and immunohistochemical protein expression. Repeated measures analysis was performed to assess the differential effect of sites on phosphorylated AKT and ERK1/2 expression variation with time. Mixed effects models were

generated to assess significance of site, time and the interaction of both factors to the expression of both proteins. All statistical tests were two-sided and p-values of 0.05 or less were considered to be statistically significant. Statistical analysis was performed with standard statistical software, including SAS Release 8.1 (SAS Institute, Cary, NC) and S-Plus 2000 (Mathsoft Inc., Seattle, WA).

Results

Detailed site- and time-dependent airway sampling of the field of injury in early stage NSCLC patients after definitive surgery

Expression patterns molecularly exemplifying the impact of smoking on the airway epithelium of cancer-free individuals have been characterized (8, 19, 29). Moreover, molecular abnormalities typically found in lung tumors have been detected in normal resected margins suggestive of a field effect (15, 25-27, 30, 31). However, the biological nature and clinical relevance of the field of injury in particular after complete removal of the tumor in early stage NSCLC patients, who are increasing in number and for whom there are no chemoprevention strategies, is yet unknown. Smoker early stage NSCLC patients were recruited on a prospective phase II surveillance clinical trial that included frequent computed tomography (CT) scans, serologies and annual bronchoscopies in which airway brushings and biopsies were obtained within one year following tumor resection and at 12, 24 and 36 months following the first time point (Supplementary methods and Figure 1). The first time point bronchoscopies were all performed within one year of definitive surgery (average, 6 months; range, 1-12). Nineteen patients were selected for the study (Table 1) based on airway sampling of at least five different sites per time point and continuously up to 24 or 36 months giving rise to 391 airway samples for

transcript-level expressing profiling. The patients were accrued between 2004 and 2008 and were comprised of former (n=16) and current (n=3) smokers. Brushings and biopsies were obtained from the main carina (MC), airways relatively adjacent (ADJ) to the previously resected tumor (ADJ) and from airways more distant from the tumor in the ipsilateral (NON-ADJ) and contralateral (CONTRA) lung (Figure 1).

Following normalization of the raw expression data, a mixed-effects model was applied to characterize the expression pattern of each gene that incorporated fixed effects such as the site from the tumor and time after surgery of the collected airway samples (Supplementary information). Histogram p-value distribution plots after fitting beta-uniform-mixture (BUM) models (32) for derived p-values based on the fixed effects suggested that both site and time of the airway samples influenced gene expression modulation (Supplementary Figures 1A and 1B).

Site-dependent differential expression patterns in airway epithelia of early stage NSCLC patients

We first sought to determine whether gene expression profiles are differentially expressed in airways by site from the tumor including those relatively adjacent to the resected tumors compared to more distant airways. Based on the generated BUM models and p-value distributions, genes differentially expressed by site were selected based on a 1% false discovery rate (FDR). We identified 1165 gene features that were statistically significantly differentially expressed by site (Supplementary Table 1). Two-dimensional hierarchical clustering demonstrated that the airway samples were divided into two main branches or clusters (Figure 2A) based on expression of the genes. Moreover, the left cluster in the indicated heat map's dendogram contained a statistically significantly higher number of adjacent airway samples

compared to the right branch which contained a significantly higher proportion of main carinas and contralateral to the tumor airways (p=0.0027 of the Fisher's exact test for count data). In addition, the two-dimensional clustering revealed eight different gene expression patterns which are indicated by the color bar and code along the left side of the heat map (Figure 2A). Some of the different gene clusters or classes were associated with a specific group of airway samples. Notably, a cluster of 263 genes (cluster 1, Supplementary Table 1), indicated by the dark green color and asterix on the heat map, was found to have highest average expression in adjacent airways (Figure 2A and Supplementary Table 2). In contrast, another cluster of 348 genes (indicated by magenta color) appeared to be highest in expression in main carinas (Figure 2A).

We then determined to functionally analyze the cluster of genes (n=263), that was found to exhibit the highest average expression in adjacent airways, between adjacent and contralateral airways. Functional pathways analysis using Ingenuity Pathways Analysis (IPA) depicted several significantly modulated pathways and molecular functions in particular, T-cell ($p=1.2 \times 10^{-9}$), chemokine C-C motif receptor 5 (CCR5) (1.5×10^{-9}) and phospholipase C signaling ($p=3.6 \times 10^{-9}$). In addition, topological gene-interaction network analysis highlighted functionally modulated and up-regulated gene networks mediated by phosphoinositide 3-kinase (PI3K) and extracellular regulate kinase (ERK) in the adjacent airways (Figure 2B). It is worthwhile to note that we also observed increased modulation of ERK MAPK-mediated gene-interaction network using a different analytical method where we compared expression profiles between adjacent and contralateral between patients in a pairwise fashion (data not shown, Supplementary sweave report 3). These findings suggest that airway site-dependent differential gene expression profiles in early stage NSCLC patients exhibit increased molecular features and gene-interaction networks typically associated with cancers.

We then determined to analyze gene expression profiles while excluding main carinas since their epithelial anatomy is suggested to be different from that of other airways (33) and thus may confound site-dependent observations. Following exclusion of main carinas, we found a reduced number of genes (n=136) that were significantly modulated by site in the field of injury (Supplementary sweave report 4). Two-dimensional hierarchical clustering demonstrated that airway samples were divided into two main branches or clusters (Supplementary Figure 2) with significantly more adjacent airways in the right branch (p=0.0002 of the Fisher's exact t-test). Moreover, the differentially expressed genes (Supplementary Table 4) were comprised of two main subgroups with one cluster (top cluster) of 113 genes exhibiting highest average expression in adjacent airways. It is important to note that when we cross-compared gene clusters that we had found to exhibit the highest average expression in adjacent airways when including (263 gene cluster) or excluding (113 gene cluster) main carinas, we found a highly significant overlap ($p=2.46 \times 10^{-191}$) in the number of genes (n=96). Moreover, the site-dependent genes identified after exclusion of main carinas were also topologically organized following functional pathways analysis into interaction networks involving PI3K and ERK.

Gene expression profiles in the lung airway epithelia of early stage NSCLC patients are modulated with time following definitive surgery

We then determined to identify genes that were differentially expressed with time. Based on the generated BUM models and p-value distributions, time-dependent differentially expressed genes were identified and selected based on a 5% FDR cut-off (n=1395, Supplementary Table 4 and Supplementary sweave report 2). Hierarchical clustering of samples indicated that the dendogram's main branches were statistically significantly unbalanced with respect to time; the

main left branch comprised a higher number of 24 and 36 month time points compared to the right cluster ($p=4.15 \times 10^{-7}$ of the Fisher's exact test for count data). Two-dimensional clustering of both genes and samples revealed two main classes of genes, those that displayed increased (upper cluster) and decreased (lower cluster) expression with time (Figure 3A). Functional pathways and gene-interaction network analysis of genes differentially expressed genes between 36 months and the first time point revealed statistically significantly modulated pathways, in particular protein ubiquitination (5.3×10^{-5}), glutathione metabolism (6.0×10^{-5}), mitochondrial dysfunction (1.4×10^{-4}) and oxidative phosphorylation (2.9×10^{-3}) as well as eukaryotic initiation factor 2 (eIF2) signaling (2.6×10^{-3}). In addition, network analysis highlighted functionally modulated and elevated gene-interaction networks with time in particular those mediated by Nuclear factor-kappa B (NF-κB), ERK, AKT and cyclin-B1 (CCNB1) (Figure 3B).

We then sought to determine the relationship of genes that were significantly modulated by site and time in the molecular field of injury. A smooth scatter plot of transformed p-values indicated that site- and time-dependent expression modulations were largely independent (Supplementary Figure 3 and Supplementary sweave report 2). We then cross-compared the site-dependent (n=1165) and time-dependent (n=1395) profiles that we had noted in the molecular field of injury. Using hypergeometric tests for overlapping probability, we found no significant overlap between genes we had determined to be differentially expressed by site and time in the molecular field of injury (p=0.865) (Supplementary sweave report 4).

Increased expression of phosphorylated AKT and ERK1/2 in airway epithelia by site from the tumor and with time following surgery

Our findings on the modulation of PI3K- and ERK1/2-mediated networks by site and time after surgery prompted us to examine the immunohistochemical (IHC) expression of surrogate markers of both signaling cascades in corresponding formalin-fixed paraffin embedded (FFPE) airway biopsies. We sought to examine expression of AKT phosphorylated at Threonine308 since phosphorylation of this amino acid is well known to occur through phosphoinositide-dependent kinase 1 (PDPK1) following PI3K activation (34). We assessed by IHC the immunohistochemical expression of phospho-AKT(Thr308) and phospho-ERK1/2(Thr202/Tyr204) in available and evaluable histologically normal bronchial epithelia biopsies (n=324) corresponding to the brushings analyzed by expression profiling. Immunoreactivity of phospho-AKT (min 0, max 300) was variable as depicted in the representative photomicrographs in Figure 4A and was detected in both the cytoplasm and nucleus of normal bronchial epithelia (NBE) (Figure 4A). IHC analysis demonstrated that cytoplasmic ($p<0.0001$) and nuclear ($p=0.01$) phospho-AKT statistically significantly increased with time up to three years in all NBE (Figure 4B) with highest expression at the 36 month time point. Nuclear phosphorylated AKT was also statistically significantly increased in adjacent NBE compared to relatively airways more distant from tumors in the mixed-effects model ($p=0.04$) (Figure 4B).

Immunoreactivity of phospho-ERK1/2 was also variable (min 0, max 209) and mainly cytoplasmic (Figure 4C). IHC analysis demonstrated that phosphorylated ERK1/2 was statistically significantly elevated in adjacent NBE ($p=0.03$) and significantly increased with time up to three years in all airways when averaged together ($p=0.02$) (Figure 4D) in the model.

Notably, there was a significant interaction (p=0.019) between site of NBE and time of sampling as phospho-ERK1/2 expression was significantly increased in adjacent NBE but not in contralateral airways and main carinas in the model (Figure 4D) with highest expression in adjacent airways and non-adjacent (ipsilateral, green plot) airways in observed at the latest time point. Similar data were obtained when we excluded main carinas in the mixed-effects model (data not shown). Moreover, we noted similar findings when all airway samples were analyzed irrespective of the presence of preneoplastic lesions (e.g. dysplasias).

These data demonstrated that like differential gene expression profiles within the lung airway field of injury, canonical activated oncogenes are modulated by site from the resected tumor and time following definitive surgery in early stage NSCLC patients.

Discussion

In this report, we characterized differential expression profiles and protein expression within the lung airway field of injury of early stage NSCLC patients by site from the tumor and time in years following surgery. We demonstrated, and to our knowledge for the first time, that gene expression profiles in histologically normal airways of early stage NSCLC patients are non-uniform by site and are modulated with time up to three years following surgery. Moreover, functional analysis of the expression profiles demonstrated that canonical expression patterns and protein kinase activation, typical of tumors, are increased in airway sites adjacent to tumors as well as remain or are further modulated in the lung airway field of injury for three years after definitive surgery. In particular, phosphorylated ERK MAPK and AKT expression were up-regulated in normal bronchial epithelia with time and by site from tumors. In light of the growing sub-population of early stage NSCLC, our findings are, in part, proof of principle and raise the

intriguing possibility of the importance of intense surveillance and molecular characterization of the remaining smoking-injured airway epithelia and its potential integration in the future into clinical practice and management of early stage disease. However, it is noteworthy that our study's patient cohort, despite its uniqueness in which expression profiling was performed on airways from multiple sites collected during bronchoscopies serially performed for 36 months following surgery, is of limited size. Moreover, the reported findings warrant the need for validation or confirmation in independent larger sets.

There is a large body of evidence that patients who have survived an upper aerodigestive cancer comprise a high risk population that may be targeted for early detection and chemoprevention efforts (5, 31). Currently there are no established adjuvant treatments in the tertiary prevention setting of early stage NSCLC patients. It has been suggested that failures in advances of chemoprevention are in part due to lack of clear and specific molecular targets (5, 35). Our extensive profiling of the molecular field of injury in NSCLC patients identified cancer-associated pathways (PI3K and ERK) aberrantly regulated in normal bronchial epithelia of NSCLC patients after tumor removal. In this context, it is plausible to suggest that a thorough characterization of the molecular field of injury in early stage NSCLC patients can aid in identification of aberrantly expressed pathways, e.g. PI3K and ERK, which could potentially serve as suitable targets for chemoprevention. However, it cannot be neglected that the alternative hypothesis can counter argue that activation of PI3K and ERK MAPK pathways may be beneficial for chemoprevention since markers of both pathways (phosphorylated AKT and ERK) increased following surgical tumor resection. Our suggestion that such pathways may serve as chemoprevention targets should be interpreted cautiously and is presented owing to the

known promalignant function of the pathways and gene networks highlighted in our analysis (36).

The first time point brushings in this study were obtained within one year of patient accrual and not at the time when the tumor was still *in situ*. The variable starting time point from surgery among patients is a limiting factor in our analysis since it is plausible to assume that the molecular field of injury may vary upon tumor removal. We were not able to avoid this caveat given the difficulty of accrual of patients following surgical tumor resection to obtain bronchial brushings at six different anatomical sites in the bronchial tree annually up to three years. However, it is important to mention that the time-dependent gene expression profiles we had identified, despite not incorporating the molecular field effect at the time the tumor was present, demonstrated gradual changes in expression with time. This effect was also evident when analyzing the immunohistochemical expression of phosphorylated AKT and ERK in corresponding bronchial biopsies with highest expression at three years. Our findings warrant future larger studies in which the molecular field of injury at the time the tumor is still *in situ* can also be serially monitored for several years.

The patient population we had studied was comprised of early stage NSCLC patients, which is in contrast to earlier transcriptomic studies of the molecular field of injury that focused on phenotypically normal smokers and non-smokers (19, 21, 22). It is still not clear whether the differences in expression described in this study reflect an already present gradient field of injury that may have contributed to tumor development in light of the differential cancer-associated pathways identified or one that arises due to the molecular impact of the tumor on the adjacent field. It is important to note that in this study, the spatial and temporal molecular field of injury in lung cancer patients was profiled prospectively starting within one year following definitive

surgery. Thus, the above speculation may be addressed by a similar thorough spatial and temporal characterization of the molecular field of injury before and after surgery in early stage patients. In addition, we did not have access to similar type of brushings from cancer-free individuals such as high-risk smokers. Similar analysis of the molecular field of injury in cancer-free individuals will shed light on the nature of site- and time-dependent expression patterns in the field of injury and whether such changes in cancer patients reflect recovery from surgery (temporal profiles) or are a cause or consequence of tumor development in the adjacent field (spatial profiles).

In conclusion, our unique study identified gene expression profiles, functional gene-networks and activated levels of oncogenic protein kinases within the field of injury of early stage NSCLC patients that are modulated or increased in airways spatially from the tumor and temporally following surgery. Moreover, the herein previously uncharacterized airway cancer-associated expression and protein kinase alterations harbor potential valuable targets for chemoprevention and warrant confirmation and further studies in larger independent cohorts.

Table 1. Clinicopathological features of NSCLC patients included in the study

Patient	Histology	Anatomical site	Stage	Recurrence	Adjuvant chemotherapy	Gender	Age	Smoking status	Pack years	Years quit smoking	Months to inclusion from surgery
1	ADC	LUL	IIA	Yes	No	Male	81	Former	100	12.0	1
2	ADC	RLL	IB	No	No	Male	64	Former	27	19.1	5
3	ADC	RUL	IB	Yes	Yes	Male	58	Former	12	25.5	7
4	SCC	RLL	IB	Yes	Yes	Female	62	Former	60	7.6	7
5	ADC	RLL	IA	No	No	Female	60	Former	31	11.8	6
6	ADC	LLL	IA	Yes	Yes	Male	62	Former	40	5.8	5
7	ADC	RUL	IB	Yes	Yes	Male	61	Current	70	NA	10
8	SCC	LUL	IB	No	No	Male	64	Current	100	NA	5
9	ADC	RUL	IA	No	No	Male	69	Former	12	36.5	8
10	ADC	RUL	IB	No	No	Female	63	Former	18	8.6	5
11	ADC	RUL	IB	Yes	No	Female	53	Former	42	5.8	3
12	SCC	RUL	IA	No	No	Male	62	Former	68	0.4	5
13	SCC	LUL	IA	No	No	Male	70	Former	96	1.4	7
14	ADC	LUL	IA	Yes	Yes	Female	45	Former	12.5	0.9	12
15	ADC	RLL	IA	No	No	Female	57	Current	80	NA	4
16	ADC	LLL	IA	No	No	Male	65	Former	48	21.0	7
17	ADC	LLL	IA	No	No	Male	64	Former	92	5.1	7
18	ADC	RML	IB	No	No	Male	71	Former	102	2.3	2
19	ADC	LUL	IA	No	No	Female	66	Former	50	18.5	5

*Anatomical site: location of primary tumor in the lung (LUL, left upper lobe; LLL, left lower lobe; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe

*Smoking status: smoking status at time of inclusion into the study (patient 7 quit smoking during the course of the study).

*Years quit smoking: Years from smoking cessation to time of inclusion into the study.

*Months to inclusion from surgery: Months elapsed from surgery to time of inclusion into the study/first bronchoscopy time point.

ADC, adenocarcinoma; SCC, squamous cell carcinoma

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin. 2011;61:69-90.

2. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA Cancer J Clin. 2012;62:10-29.

3. Herbst RS, Heymach JV, Lippman SM. Lung cancer. N Engl J Med. 2008;359:1367-80.

4. Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med.365:395-409.

5. Gold KA, Kim ES, Lee JJ, Wistuba, II, Farhangfar CJ, Hong WK. The BATTLE to personalize lung cancer prevention through reverse migration. Cancer Prev Res (Phila).4:962-72.

6. Goldstraw P, Ball D, Jett JR, Le Chevalier T, Lim E, Nicholson AG, et al. Non-small-cell lung cancer. Lancet. 2011;378:1727-40.

7. Gazdar AF, Thun MJ. Lung cancer, smoke exposure, and sex. J Clin Oncol. 2007;25:469-71.

8. Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. Cancer Prev Res (Phila Pa). 2008;1:396-403.

9. Auerbach O, Stout AP, Hammond EC, Garfinkel L. Changes in bronchial epithelium in relation to cigarette smoking and in relation to lung cancer. N Engl J Med. 1961;265:253-67.

10. Mao L, Lee JS, Kurie JM, Fan YH, Lippman SM, Lee JJ, et al. Clonal genetic alterations in the lungs of current and former smokers. J Natl Cancer Inst. 1997;89:857-62.

11. Powell CA, Klares S, O'Connor G, Brody JS. Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer. Clin Cancer Res. 1999;5:2025-34.

12. Wistuba, II, Lam S, Behrens C, Virmani AK, Fong KM, LeRiche J, et al. Molecular damage in the bronchial epithelium of current and former smokers. J Natl Cancer Inst. 1997;89:1366-73.

13. Franklin WA, Gazdar AF, Haney J, Wistuba, II, La Rosa FG, Kennedy T, et al. Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. J Clin Invest. 1997;100:2133-7.

14. Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, et al. Aberrant methylation of p16(INK4a) is an early event in lung cancer and a potential biomarker for early diagnosis. Proc Natl Acad Sci U S A. 1998;95:11891-6.

15. Belinsky SA, Palmisano WA, Gilliland FD, Crooks LA, Divine KK, Winters SA, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res. 2002;62:2370-7.

16. Heller G, Zielinski CC, Zochbauer-Muller S. Lung cancer: from single-gene methylation to methylome profiling. Cancer Metastasis Rev.29:95-107.

17. Soria JC, Rodriguez M, Liu DD, Lee JJ, Hong WK, Mao L. Aberrant promoter methylation of multiple genes in bronchial brush samples from former cigarette smokers. Cancer Res. 2002;62:351-5.

18. Zochbauer-Muller S, Lam S, Toyooka S, Virmani AK, Toyooka KO, Seidl S, et al. Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. Int J Cancer. 2003;107:612-6.

19. Spira A, Beane J, Shah V, Liu G, Schembri F, Yang X, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. Proc Natl Acad Sci U S A. 2004;101:10143-8.

20. Schembri F, Sridhar S, Perdomo C, Gustafson AM, Zhang X, Ergun A, et al. MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. Proc Natl Acad Sci U S A. 2009;106:2319-24.

21. Spira A, Beane JE, Shah V, Steiling K, Liu G, Schembri F, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nat Med. 2007;13:361-6.

22. Gustafson AM, Soldi R, Anderlind C, Scholand MB, Qian J, Zhang X, et al. Airway PI3K pathway activation is an early and reversible event in lung cancer development. Sci Transl Med.2:26ra5.

23. Wistuba, II, Behrens C, Virmani AK, Mele G, Milchgrub S, Girard L, et al. High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. Cancer Res. 2000;60:1949-60.

24. Wistuba, II, Behrens C, Virmani AK, Milchgrub S, Syed S, Lam S, et al. Allelic losses at chromosome 8p21-23 are early and frequent events in the pathogenesis of lung cancer. Cancer Res. 1999;59:1973-9.

25. Nelson MA, Wymer J, Clements N, Jr. Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers. Cancer Lett. 1996;103:115-21.

26. Tang X, Shigematsu H, Bekele BN, Roth JA, Minna JD, Hong WK, et al. EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. Cancer Res. 2005;65:7568-72.

27. Tang X, Varella-Garcia M, Xavier AC, Massarelli E, Ozburn N, Moran C, et al. Epidermal growth factor receptor abnormalities in the pathogenesis and progression of lung adenocarcinomas. Cancer Prev Res (Phila Pa). 2008;1:192-200.

28. Sun M, Behrens C, Feng L, Ozburn N, Tang X, Yin G, et al. HER family receptor abnormalities in lung cancer brain metastases and corresponding primary tumors. Clin Cancer Res. 2009;15:4829-37.

29. Beane J, Sebastiani P, Liu G, Brody JS, Lenburg ME, Spira A. Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. Genome Biol. 2007;8:R201.

30. Wistuba, II, Behrens C, Milchgrub S, Bryant D, Hung J, Minna JD, et al. Sequential molecular abnormalities are involved in the multistage development of squamous cell lung carcinoma. Oncogene. 1999;18:643-50.

31. Wistuba, II, Gazdar AF. Lung cancer preneoplasia. Annu Rev Pathol. 2006;1:331-48.

32. Pounds S, Morris SW. Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. Bioinformatics. 2003;19:1236-42.

33. Lips KS, Volk C, Schmitt BM, Pfeil U, Arndt P, Miska D, et al. Polyspecific cation transporters mediate luminal release of acetylcholine from bronchial epithelium. Am J Respir Cell Mol Biol. 2005;33:79-88.

34. Engelman JA. Targeting PI3K signalling in cancer: opportunities, challenges and limitations. Nat Rev Cancer. 2009;9:550-62.

35. Khuri FR. The dawn of a revolution in personalized lung cancer prevention. Cancer Prev Res (Phila). 2011;4:949-53.

36. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646-74.

Figure legends

Figure 1. Spatial and temporal molecular field of injury in early stage NSCLC patients after definitive surgery. Schematic depicting the site- (top) and time- (bottom) dependent collection of airway epithelia brushings by bronchoscopy. Smoker early-stage NSCLC patients were enrolled into a surveillance clinical trial for annual follow-up and bronchoscopies within one year after definitive surgery. Bronchial airway epithelial cells (brushings and biopsies) were obtained from five to six different sites comprised of the main carina (MC), 4 airways from 4 lobes ipsilateral (NON-ADJ) or contralateral (CONTRA) to the originally resected tumor and of the bronchial stump or main stem bronchus adjacent to the tumor and lobe. All site-different airway epithelia were collected at the time of inclusion in the study and at 12, 24 and 36 months following the starting time point (391 airways from 19 NSCLC patients).

Figure 2. Site-dependent airway epithelia differential gene expression patterns. A. Heat map depicting two-dimensional clustering of airway samples (n=391) and genes (n=1165) that were determined to be differentially expressed by site in the mixed-effects model based on a 1% FDR cut-off. The identified eight gene clusters are labeled with different colors with the green cluster of genes (n=263, *A) exhibiting highest average expression in adjacent airways and the magenta cluster (n=348, *MC) having highest expression in main carinas. **B**. Gene-interaction analysis by IPA depicting networks with significant scores that indicate the likelihood of genes in a network being found together than due to chance. The depicted networks were found to be mediated by PI3K (top) and ERK1/2 (bottom) with both kinases themselves not modulated in expression. Gene expression variation based on the statistical cut-off described above is depicted by color in the network (red, up-regulated; green, down-regulated).

Figure 3. Temporal modulation of the molecular field of injury after definitive surgery in early stage NSCLC patients. A. Heat map depicting two-dimensional clustering of airway samples (n=391) and genes (n=1395) that were determined to be differentially expressed by time in the mixed-effects model based on a 5% FDR cut-off. **B**. Gene-interaction analysis, similar to that in Figure 2, by IPA depicting networks with increased likelihood of genes being found together than due to chance and mediated by NF-κB, ERK1/2, AKT and CCND1. CCND1 itself was up-regulated at the expression level. Gene expression variation based on the statistical cut-off is depicted by color in the network (red, up-regulated; green, down-regulated).

Figure 4. Site- and time-dependent immunohistochemical expression of phosphorylated AKT and ERK1/2 in the airway field of injury. A. Representative photomicrographs (20x magnification) depicting strong (top) and weak (bottom) phospho-AKT(Thr308) immunostaining. **B.** Immunohistochemical scores of cytoplasmic (left) and nuclear phospho-AKT (middle) were assessed for statistically significant differences by site and time in a mixed-effect model and plotted in main carinas (MC) and in adjacent (ADJ), non-adjacent (ipsilateral, NON-ADJ) and contralateral (CONTRA) airways with time. **C.** Representative photomicrographs (20x magnification) depicting strong (top) and weak (bottom) phospho-ERK1/2(Thr202/Tyr204) immunostaining. **D.** Immunohistochemical scores of phosphorylated ERK1/2 levels were assessed for statistically significant differences by site and time in a mixed-effect model and repeated measures analysis, site*time, term for interaction between site and time factors. Error bars represent S.E. of the mean.

Figure 1

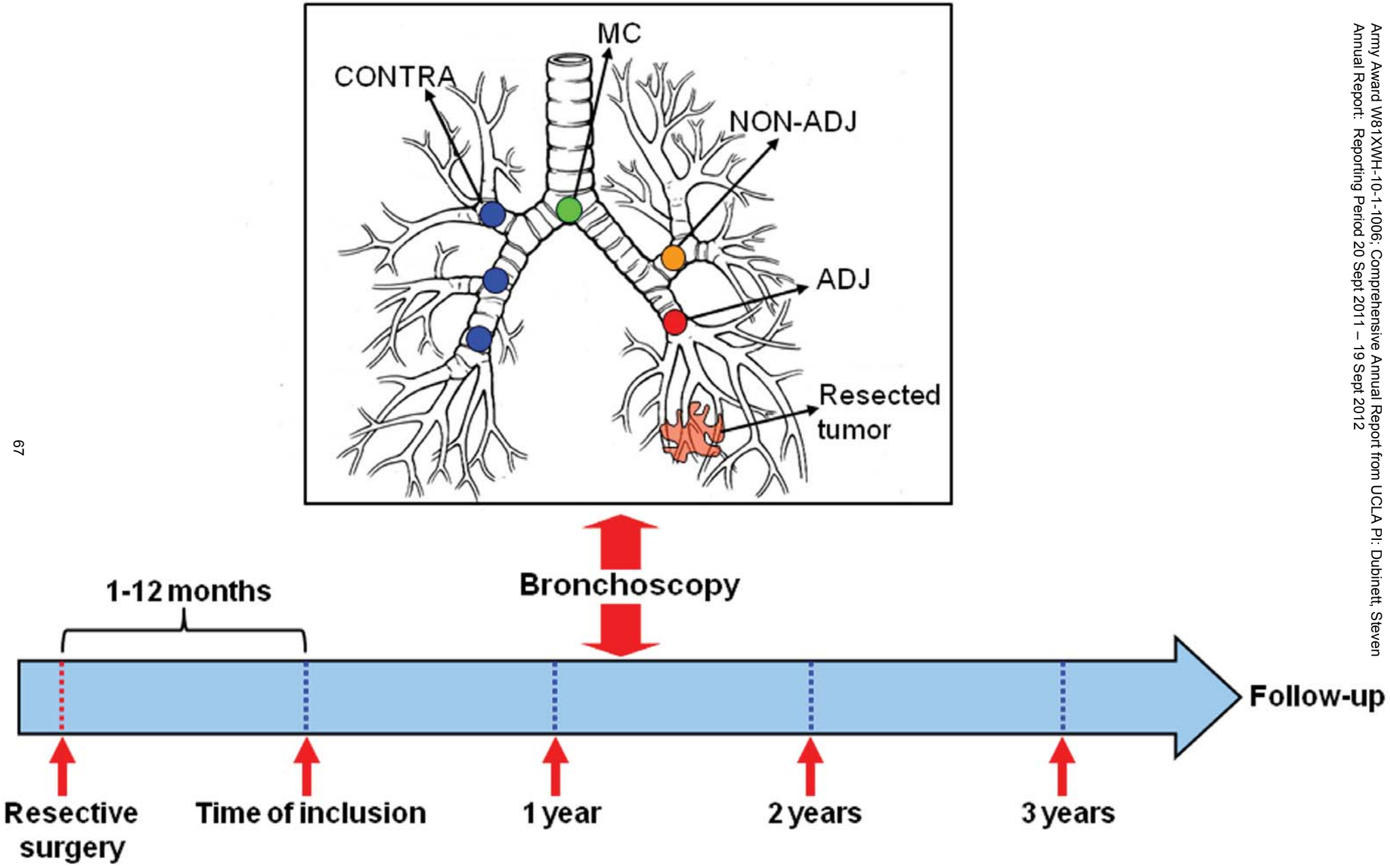

MC
CONTRA
NON-ADJ
ADJ
Resected tumor
Bronchoscopy
1-12 months
Follow-up
Resective surgery
Time of inclusion
1 year
2 years
3 years

Figure 2

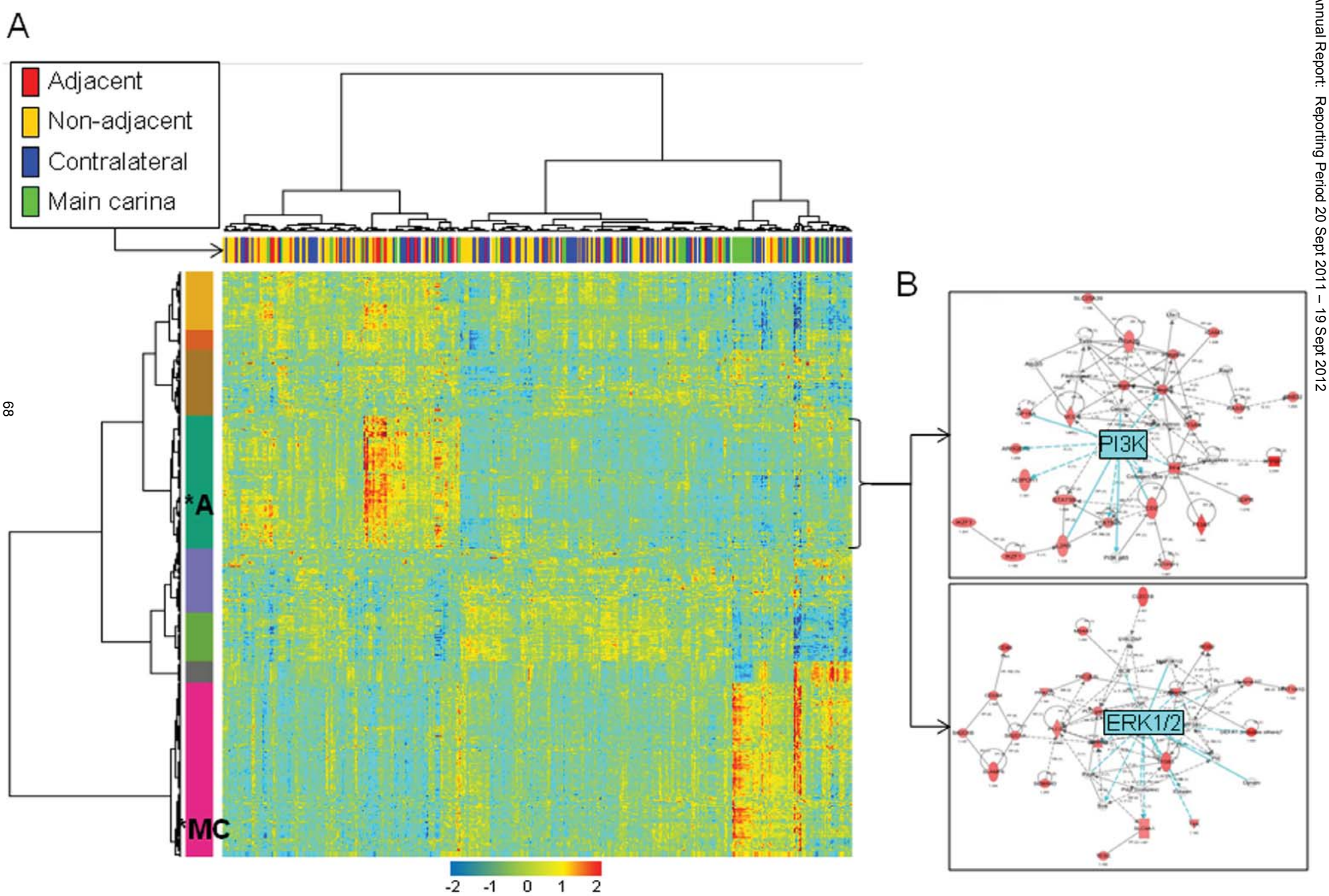
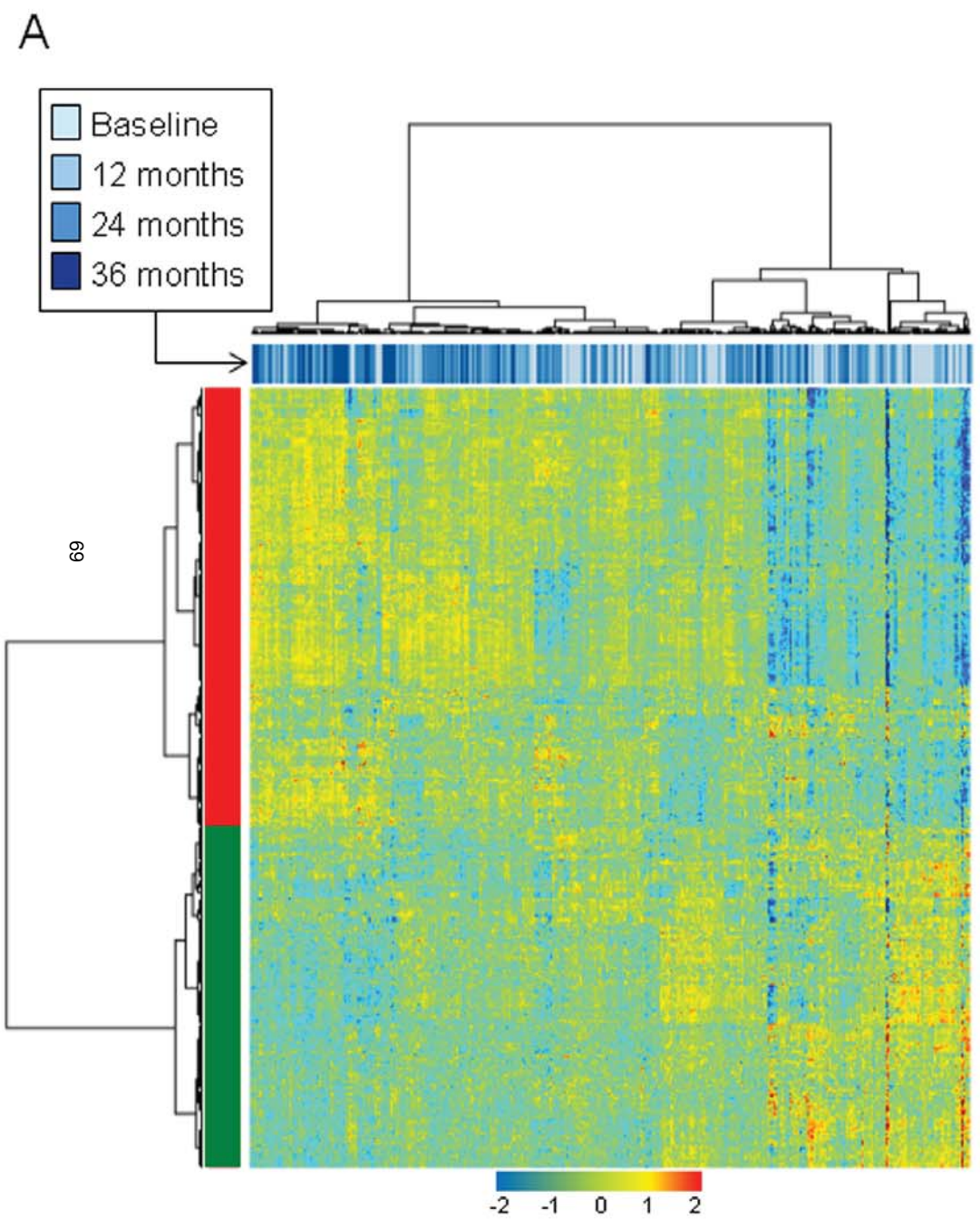

A
Adjacent
Non-adjacent
Contralateral
Main carina
*A
*MC
-2 -1 0 1 2
B
PI3K
ERK1/2

Figure 3


A
Baseline
12 months
24 months
36 months
-2
-1
0
1
2

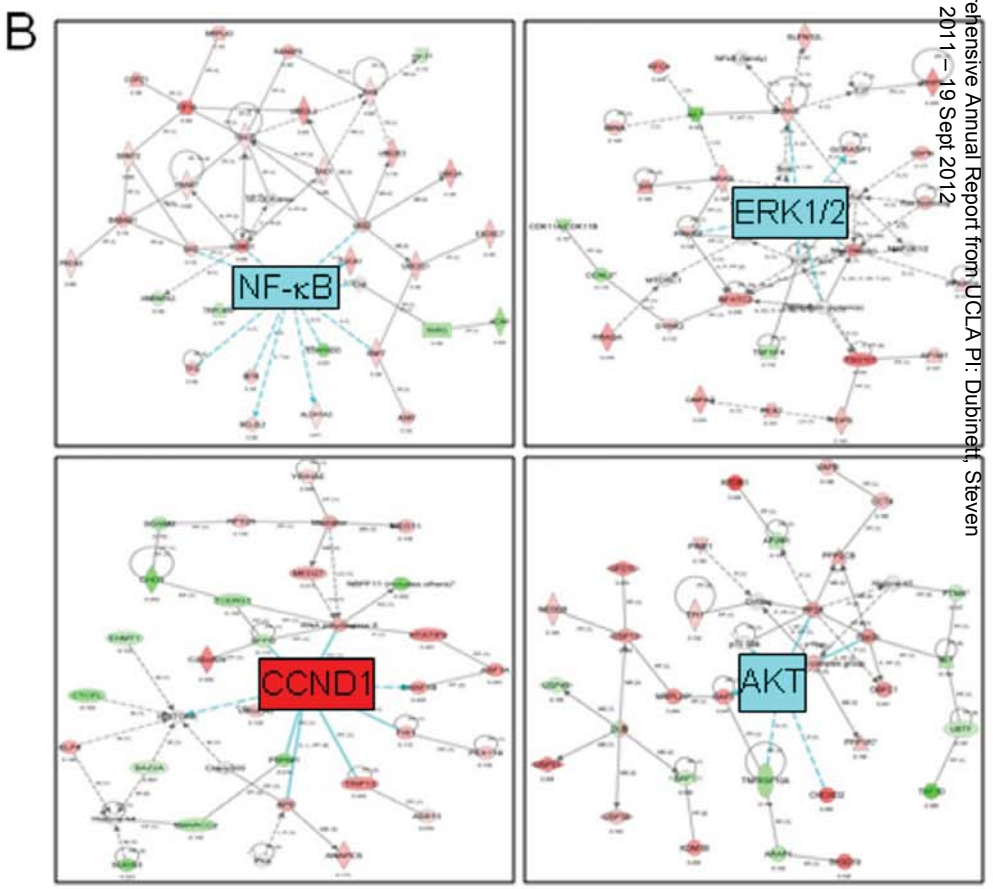
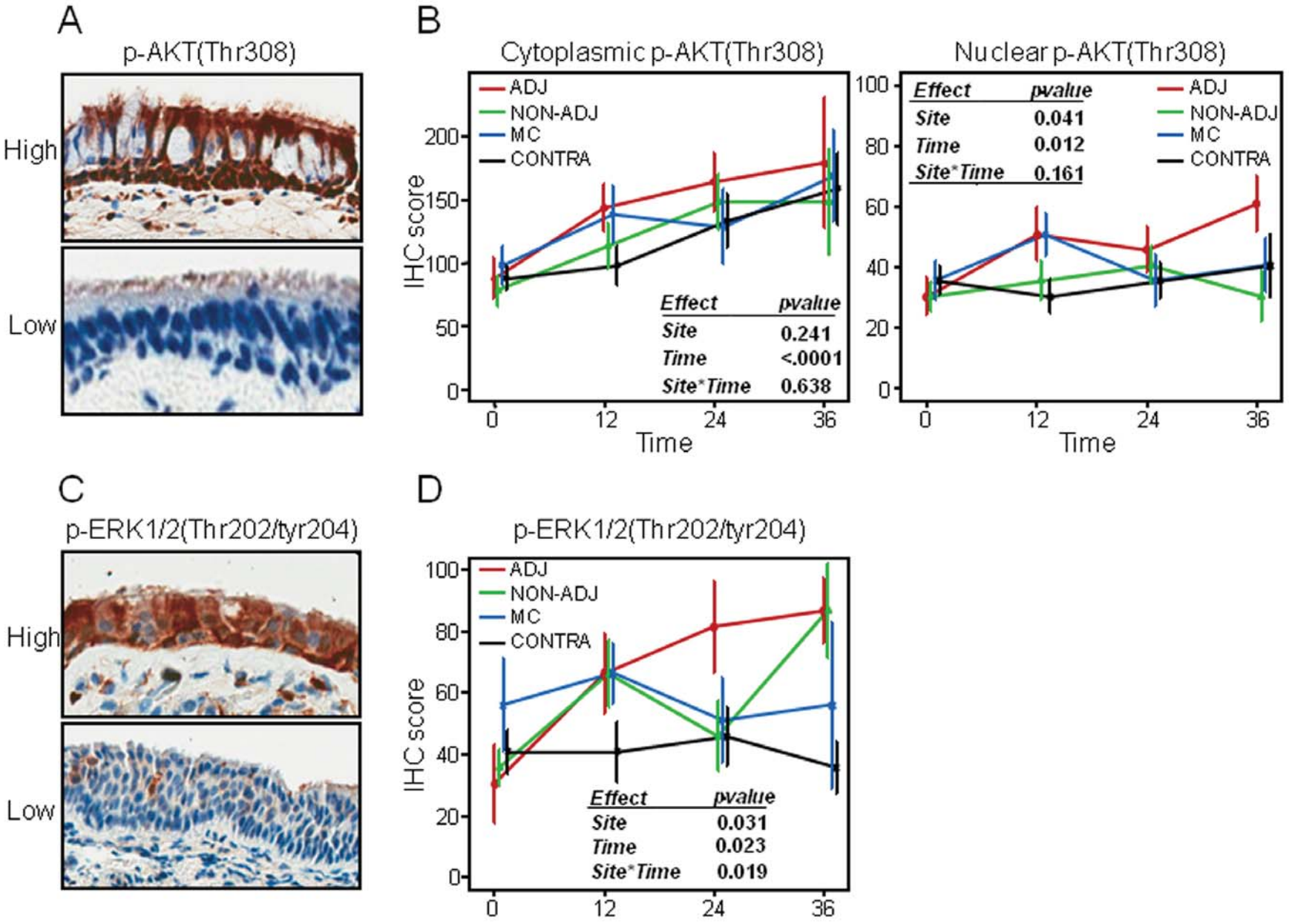

B
NF-κB
ERK1/2
CCND1
AKT

Figure 4


A
p-AKT(Thr308)
High
Low
B
Cytoplasmic p-AKT(Thr308)
ADJ
NON-ADJ
MC
CONTRA
IHC score
Time
Effect pvalue
Site 0.241
Time <.0001
Site*Time 0.638
Nuclear p-AKT(Thr308)
Effect pvalue
Site 0.041
Time 0.012
Site*Time 0.161
C
p-ERK1/2(Thr202/tyr204)
High
Low
D
p-ERK1/2(Thr202/tyr204)
Effect pvalue
Site 0.031
Time 0.023
Site*Time 0.019